

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 12, 2005, 10:30:41 ; Search time 10000 Seconds  
(without alignments)  
5194.399 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPTGVLSLLLVITAGC.....VSMNGSIRNGASFVSCSKDR 1072

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0		0.5	
Ygapop 6.0		7.0	
Delop 6.0		7.0	

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cn2.1/USPTO.spool.p/US10764390/runat.12102005.110203.22329/app.query.fasta_1.1223
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODES=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10764390@CGN 1 1 6626 @runat.12102005.110203.22329 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sv.\*

13: gb.un.\*

14: gb.vi.\*

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5574	99.9	6791	9 AB002317	AB002317 Homo sapi
2	4255.5	76.3	4971	10 AK122246	AK122246 Mus muscu
3	4092	73.3	2433	6 CQ726592	CQ726592 Sequence
4	2435.5	43.6	4379	10 BC028869	BC028869 Mus muscu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	2405	43.1	4183	9	BC014530	BC014530 Homo sapi
6	2404	43.1	3908	6	AX195570	AX195570 Sequence
7	2404	43.1	4000	9	AY163234	AY163234 Homo sapi
8	2385	42.7	4001	9	AF275679	AF275679 Homo sapi
9	2269	40.7	2970	6	CQ727300	CQ727300 Sequence
10	2184.5	39.1	2796	6	BD160555	BD160555 Primer fo
11	2184.5	39.1	2796	6	AX883831	AX883831 Sequence
12	2184.5	39.1	2796	9	AK024287	AK024287 Homo sapi
13	2122.5	38.0	2743	9	BC031672	BC031672 Homo sapi
14	2005.5	35.9	4423	9	AB058740	AB058740 Homo sapi
15	1817.5	32.6	2501	6	BD057820	BD057820 Secreted
16	1798	32.2	5555	6	BD183361	BD183361 Novel gen
17	1675.5	30.0	2811	10	BC022154	BC022154 Mus muscu
18	1451.5	26.0	3345	6	CQ589125	CQ589125 Sequence
19	1443.5	25.9	2223	6	BD157975	BD157975 Primer fo
20	1443.5	25.9	2223	6	AX879698	AX879698 Sequence
21	1443.5	25.9	2223	9	AK022734	AK022734 Homo sapi
22	1426	25.6	3727	3	BT010098	BT010098 Drosophil
23	1395	25.0	2322	9	AK090878	AK090878 Homo sapi
24	1344	24.1	28692	9	AL512385	AL512385 Human DNA
25	1280.5	22.9	5775	6	CQ589124	CQ589124 Sequence
26	1280.5	22.9	94236	2	AC020539	AC020539 Drosophil
27	1280.5	22.9	185790	3	AC091221	AC091221 Drosophil
28	1280.5	22.9	187555	3	AC010037	AC010037 Drosophil
29	1280.5	22.9	281917	3	AE003556	AE003556 Drosophil
30	1274.5	22.8	3133	3	AK114825	AK114825 Ciona int
31	1267.5	22.7	3121	3	AK113004	AK113004 Ciona int
32	1166	20.9	57708	2	AC130340	AC130340 Homo sapi
33	996	17.8	1000	5	CR353665	CR353665 Gallus ga
34	931.5	16.7	1874	9	BC007645	BC007645 Homo sapi
35	886.5	15.9	256608	10	AL589699	AL589699 Mouse DNA
36	880.5	15.8	245394	2	AC125756	AC125756 Rattus no
37	845	15.1	772	6	BD148725	BD148725 Primer fo
38	845	15.1	772	6	AX868663	AX868663 Sequence
39	823	14.7	818	6	AX677257	AX677257 Sequence
40	786	14.1	57349	2	AC135073	AC135073 Homo sapi
41	734	13.2	720	5	EX933224	EX933224 Gallus ga
42	681	12.2	580	6	BD071994	BD071994 Secreted
43	640	11.5	578	6	BD071762	BD071762 Secreted
44	525.5	9.4	67498	2	AC012648	AC012648 Drosophil
45	506.5	9.1	46913	2	AC023733	AC023733 Drosophil

#### ALIGNMENTS

RESULT 1	AB002317	Homo sapiens mRNA for KIAA0319 gene, partial cds.	6791 bp	mRNA	linear	PRI 10-JAN-2004
LOCUS	AB002317					
DEFINITION	AB002317					
ACCESSION	AB002317					
VERSION	AB002317.1					
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						

[illegible]



3'-end one pass sequencing.	
FEATURES	Location/Qualifiers
source	1..4971
	/organism="Mus musculus"
	/mol_type="mRNA"
	/db_xref="taxon:10090"
	/clone="mbg09941"
	/tissue_type="brain"
	/dev_stage="adult"
	/note="vector:modified pBC SK+"
gene	1..4971
	/gene="mKIAA0319"
CDS	<249..3500
	/gene="mKIAA0319"
	/note="CDS is predicted by in silico analysis. Start codon is not identified."
	/evidence=not_experimental
	/product="mKIAA0319 protein"
	/protein_id="BAC65528.1"
	/db_xref="GI:28972149"
	/translation="GRMVSPRGVLSLLLAAMAGSSQCSGRTYSDAIIISPNETIRIMVSTFSGDCTAACDDLLTCDLAWFFGSCVLYKMRSENCEPRTTGPIRSYLYFVRPVQRPGLLDYGMMLSKGSPSGAWGSDLEDRLKDLPLGKGGPBEETIYSDFYKDLERGLQPSNQDPRGSABYPDMSLLPSNCGFNATATGDNASAMEKLQDPTPHLDQRLQALANESTWSPTPGHSSISSVWPSSASPLPTEGLEGBETLQLOQPNSNSKEVPMPNSPASPLESPPATTEKNSNFTVTPRSKHSPTPTFTSVLTGLTPPPWP LPTASRTWKALAVSAGDNILVLTLPDREAEKASVEPAPADTVSYEWSLMSHPVDFQGIKQENKPLHLSQSLVGLYAFRAVAVSENAFGEYGVNVTMPAARVNPQPVAVVSFOQLSVFLTSALIDQSQDDEIYVSHVEVDGFFLGEFFPADTILRLSNLVPGNVTFRLTIDSGATNSTASLIVIRDAVDYPPVANAGPNTITLPTNTIILNQSSDDHOIVLVYFWFAGGGSKEMVQGAOTPYLHLSQEGEYTPQLMVTDSGSGQSTALVAVTQAEENQAPVAVAPDKELVPVQSATLDGSRSSDDHGVICVYHWEHIRGFSAVEMENVDAIATVTGLQVGIYFRLTYRQOGLSSTSLTVAVKKNSSPPRAOAGGRVILILPNSITLDGSDTDRGIVSYLWIRDGSPAGADVITGSDHRAALQTLNIVEGYITFHLVTSQGSADSDAVEPLPKKDGVELILQVGEQTEQOKETLVRLQALVLLNVLDSVTKIQIQAHTDVTSTVIVFYVQSGSPFKVLRAAAARVNLHKLRSKEAEFLF KVLVRVTAGVLSKCHGHCDPITKRCICSQLMENLIQRYWMDGSENCWGFVFAA LALTLLTAGVLSMCLICCCRRKRTKIRKTKYTILDSMDRQERWELAPKYGIKHSRSTENSLMVSSEFSDQDTLFSRERMERGVKLSLNGCARNGVSPGYYSKDR"
ORIGIN	
Alignment Scores:	
Pred. No.:	3,12e-220 Length: 4971
Score:	4255.50 Matches: 820
Percent Similarity:	84.20% Conservative: 91
Best Local Similarity:	75.79% Mismatches: 160
Query Match:	76.26% Indels: 11
DB:	10 Gaps: 4
US-10-764-390-3 (1-1072) x AK122246 (1-4971)	
Qy	1 MetAlaProThrGlyValLeuSerSerLeuLeuLeuValThrIleAlaGlyCys 20
Db	255 ATGTGTCTCCACACGAGTACTGTCTTCAATGCTGTCTGGCAGCCATGGCAGGAGGC 314
Qy	21 AlaArgLysGlnCysSerGluGlyArgCysThrSerAsnAlaValIleSerProAsnLeu 40
Db	315 AGTTCTCAGCAGTGCTCTGAGGGCAGGACTTACTCCGATGCCATCTTCCACCTAACCCG 374
Qy	41 GluThrThrArgIleMetArgValSerHisThrPheProValValAlaAspCysThrAlaAla 60
Db	375 GAAACCATCAGATCATGCGGGTGTCTCAAAACCTTCTCCGTGGGAGACTGCACGGCGCT 434
Qy	61 CysCysAspLeuSerSerCysAspLeuAlaTrpTrpPheGluGlyArgCysTyrLeuVal 80
Db	435 TGCTGTGACTTGCTCACCTGTGACCTGCGCTGGTGGTTTTCAGGGCAGCTGCTATCTGGTG 494
Qy	81 SerCysProHisLysGluAsnGlyGluProLysLysMetGlyProIleArgSerTyrLeu 100
Db	495 AATATGATGGCTCGAGAGATTGGAGCCACGACACACAGGCCATCCCGTCTTACCTC 554
Qy	101 ThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMet 120
Db	555 ACTTTCGTGCGCAGACCTGTCCAGAGGCCCGGCGAGCTGTGGACTATGGAGACATGATG 614
Qy	121 LeuAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArgLysAsp 140
Db	615 CTGAGCAGGGGCTCCCTCCCTCAGGAGCTTGGGGAGACTCCCTTGGAGACTCAGGAAGAC 674
Qy	141 Leu***PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTyr***AspAspTyr 160
Db	675 TTGCCCTTCTTGGCAAGACGGGGACACGAGGAGACCACTGAGTACTCAGATGAATAC 734
Qy	161 ArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAla 180
Db	735 AAAGACCTTGGAGCGGGCTCTCTGCAGCCCTCCAAACAGCAGACCCCTAGAGGTAGCGCA 794
Qy	181 GluTyrThrAspTrpGlyLeuProGlySerGluGlyAlaPheAsn---SerSerVal 199
Db	795 GAGTACCCCGGACTGGAGCTGCTGCCAGCAACAGAGGAGGTTTCAATGCTACACTACA 854
Qy	200 GlyAspSerProAlaValProAlaGluThrGlnGlnAspPro----- 213
Db	855 GGAGACAACCTCAGCTGTCTCCATGGAGAAGCTGCAGGATCCACGCCCCACCCACTGGAC 914
Qy	214 -----GluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeuProGlu 231
Db	915 CAGGAGCAGCTGCAGGCCCTGAAATGAGTCGACTTGGTCCCTTACACCCCGGCACCTCTTCA 974
Qy	232 ArgSerValLeuLeuProLeuProThrPro---SerSerGlyGluValLeuGluLys 250
Db	975 ATAAAGCATGTGTGGCCCTTCTCTCGAGGCCATTTACCTACAGAGGAGGACTGGAAGA 1034
Qy	251 GluLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLysGluValLeuMet 270
Db	1035 GAAGAGACTTTACAGCTCCAGACAACCGAGCAATAGCTCTGGAAGAGAGGTTCCAAATG 1094
Qy	271 ProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGluLysSer 290
Db	1095 CTTTCCCATATCTCTCCCTGCCAGCCTGGAGTCTAGCCACGACGACGCGGAAAAAAC 1154
Qy	291 ProValLeuThrValThrProGlySerThrGluHisSerIleProThrProProThrSer 310
Db	1155 TCCAACTTTACAGTCACGCCACGGAGCAGAAAAACACAGACCCCACTTTTCTACCACT 1214
Qy	311 AlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaProArg 330
Db	1215 ACAGTCTCTCAGTGGGCTAACCCCTCTCCGTGGCCCTTGTCACT---ACTGTTCCAGG 1271
Qy	331 ThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAsn 350
Db	1272 ACAGTAAAGCGCTTGCTGTGTCTGGAGATAACCTAGTACTTAACCTTACCGGACAGA 1331
Qy	351 GluValGluLeuLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTyr 370
Db	1332 GAAGCAGAACTGAAGGCTCTGTGAAACGAGCGCCCTTGCAGATACAACTTACTCTCAT 1391
Qy	371 GluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLys 390
Db	1392 GAATGGAGTTAATGAGCCACCCAGTAGACTTCCAGGTAAATAACAAACAAGAAAAACAG 1451
Qy	391 GlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSer 410
Db	1452 CCGACTCTTCACTCTCTCAATATCTGTGGGACTCTATGCTTCCCTTCCAGAGTGGCTGTCT 1511
Qy	411 SerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgArgVal 430
Db	1512 AGTGAGAAATGCATTTGGAGAGGCTATGTCAATGTCACGGTTATGCCAGCTCAAGAGTC 1571
Qy	431 AsnLeuProProValAlaValValSerProGlnLeuGlnLeuThrLeuProLeuThr 450
Db	1572 NACCAGCCACCTGTAGCTGTCTTCTCCCGACAGACAGAGAGCTCAGTGTGCTTTGACC 1631
Qy	451 SerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTrp 470





/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 8.54e-212 Length: 2433  
Score: 4092.00 Matches: 807  
Percent Similarity: 80.34% Conservative: 2  
Best Local Similarity: 80.14% Mismatches: 2  
Query Match: 73.33% Indels: 196  
DB: 6 Gaps: 5

US-10-764-390-3 (1-1072) x CQ726592 (1-2433)

Qy 46 MetArgValSerHisThrPheProValValAspCysThrAlaAlaCysCysAspLeuSer 65  
Db 1 ATCGGGGTGTCTCACACCTTCCTCGTAGACTGCGAGCGCGCTTGCTGACCTGTCC 60  
Qy 66 SerCysAspLeuAlaTrpPheGluGlyArgCysTyrLeuValSerCysProHisLys 85  
Db 61 AGCTGTGACCTGGCCTGGTGTTCGAGGGCGCTGCTACCTGGTGAGCTGCCCCACAAA 120  
Qy 86 GluAsnCysGluProLysLysMetGlyProIleArgSerTyrLeuThrPheValLeuArg 105  
Db 121 GAGAACTGTGAGCCCAAGAGATGGGCCCCCATCAGGTCTTATCTCACTTTTGTGCTCCGG 180  
Qy 106 ProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgGlySer 125  
Db 181 CCGTTTCAGAGGCTGCACAGCTGCTGGACTATGGGGACATGATGCTGAACAGGGGCTCC 240  
Qy 126 ProSerGlyIleTrpGlyAspSerProGluAspIleArgLysAspLeu\*\*\*PheLeuGly 145  
Db 241 CCCTCGGGGATCTGGGGGACTCACCTGAGGATATCAGAAAGGACTTGGCCCTTCTAGGC 300  
Qy 146 LysAspTrpGlyLeuGluGluMetSerGluTyr\*\*\*AspAspTyrArgGluLeuGluLys 165  
Db 301 AAAGATGGGGCTTAGAGGAGATGCTCAGTACTCAGATGACTACCGGGAGCTGGAGAAG 360  
Qy 166 AspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAlaGluTyrThrAspTrp 185  
Db 361 GAGCTCTTGAACCCCACTGAGTGGCAAGCAGAGGCCAGAGGGAGTCCCGAGTACAGGACTGG 420  
Qy 186 GlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaVal 205  
Db 421 GGCTACTGCGGGCAGCAGGGGGCTTCAACTCTCTCTTGGAGACAGTCTCTGGGTG 480  
Qy 206 ProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrPro 225  
Db 481 CCAGCGGAGACGCGAGCAGGACCTGAGCTCCATTACCTGAATGAGTGGCTTCAACCCCT 540  
Qy 226 AlaProLysLeuProGluArgSerValLeuLeuLeuProLeuProThrThrProSerGly 245  
Db 541 GCCCCAAAATCTCCCTGAGAGAAGTGTGTCTTCCCTTGGCCGACTACTCCATCTTCAGGA 600  
Qy 246 GluValLeuGluLysGluLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGly 265  
Db 601 GAGGTGTGGAGAAAGAAAGGCTTCTCAGCTCCAGGAACAATCCAGCAACAGCTCTGGA 660  
Qy 266 LysGluValLeuMetProSerHisSerLeuProProAlaSerLeuGluLeuSerSerVal 285  
Db 661 AAAGAGGTCTTAATGCCCTTCCCATGATCTTCTCCCGCAAGCCTGGAGCTCAGTCCAGTC 720  
Qy 286 ThrValGluLysSerProValLeuThrValThrProGlySerThrGluHisSerIlePro 305  
Db 721 ACCGTGGAGAAAGCCAGTGTCTCACAGTCAACCCGGGGAGTACAGAGCACAGCATCCCA 780  
Qy 306 ThrProProThrSerAlaAlaProSerGluSerThrProSerGluLeuProIleSerPro 325  
Db 781 ACACCTCCCACTAGCGCAGCCCCCTCTGAGTCCACCCCATCTGAGCTACCCATATCTCTCT 840  
Qy 326 ThrThrAlaProArgThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIlelle 345

Db 841 ACCACTGCTCCAGGACA----- 858  
Qy 346 ThrLeuProAspAsnGluValGluLeuLysAlaPheValAlaProAlaProValGlu 365  
Db 859 -----GAA 861  
Qy 366 ThrThrTyrAsnTyrGluTrpAsnLeuLeuSerHisProThrAspTyrGlnGlyGluIle 385  
Db 862 ACAACCTACAACCTATGAATGAATTTAATAAGCCACCCACACAGACTACCAAGGTGAAATA 921  
Qy 386 LysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPhe 405  
Db 922 AAACAAGACACAAGCAAACTCTTAACCTCTCTCAATTTGCGTCCGACTTTATGCTCTC 981  
Qy 406 LysValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLys 425  
Db 982 AAAGTCACTGTTTCTAGTGAACCGCTTTGGAGNAGGATTTGTCAATGTCACTGTAAAG 1041  
Qy 426 ProAlaArgArgValAsnLeuProProValAlaValValSerProGlnLeuGlnLeu 445  
Db 1042 CCT----- 1044  
Qy 446 ThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThrThrAspAspThrGluIle 465  
Db 1045 -----GAAAGTACAGATGATACTGAAATA 1068  
Qy 466 ValSerTyrHisTrpGluGluIleAsnGlyProPheIleGluGluLysThrSerValAsp 485  
Db 1069 GTGAGTTATCATTTGGGAAGAAATAAACCGGCCCTTCATAGAAGAGAACATTCAGTTGAC 1128  
Qy 486 SerProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrVal 505  
Db 1129 TCTCCGCTGTACGCTTGTCTAACCTTGATCT----- 1161  
Qy 506 ThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuIleValAsnAsnAlaVal 525  
Db 1162 -----GCCCTAATAGTGAACAAATGCTGTG 1185  
Qy 526 AspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSer 545  
Db 1186 GACTACCCACCGTTGCTTANTGCAGGACCAATATCACCATTAATTTGCCCCAAAACCTCC 1245  
Qy 546 IleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuTyrGluTrpSer 565  
Db 1246 ATCACTTTGAATGGAAACACAGAGCAGTACGATCACAGATTTGCTCTTATGAGTGTCC 1305  
Qy 566 LeuGlyProGlySerGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeu 585  
Db 1306 CTGGGTCTCTGGG----- 1317  
Qy 586 HisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSer 605  
Db 1317 ----- 1317  
Qy 606 ArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProPro 625  
Db 1318 -----AAAAACAATAGACTCCA 1335  
Qy 626 ValAlaValAlaGlyProAspLysGluLeuIlePheProValGluSerAlaThrLeuAsp 645  
Db 1336 GTGGCTGTGGCGCGCCCTGATAAAGAGCTGATCTTCCAGTGGAAAGTCTACCTCCGAT 1395  
Qy 646 GlySerSerSerSerAspAspHisGlyIleValPheTyrHisTrpGluHisValArgGly 665  
Db 1396 GGGAGCAGCAGCAGCAGTACCCAGCATTTGTTCTACCACTGGGAGCAGCTCAGAGGC 1455  
Qy 666 ProSerAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGln 685  
Db 1456 CCCAGTGCAGTGCAGATGGAAATATTGACAAAGCAATAGCCACTGTGTGCTCTCCAG 1515  
Qy 686 ValGlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerThrThrSer 705  
Db 1516 GTGGGACCTACACTTCGCTTTGACATGTAAGAACACAGGAGCTGAGCAGCAGCTCC 1575

QY 706 ThrLeuThrValAlaValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyGly 725  
 DB 1576 ACCCTCACTGGCTGTGAAGAAGAAATAATAGTCTCCAGAGCCCGGCTGGTGC 1635  
 QY 726 ArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAsp 745  
 DB 1636 AGACATGTTCTTGCTCCCAATATTCATTTGGATGGTTCAAGGTCTACTGAT 1695  
 QY 746 AspGlnArgIleValSerTyrLeuTyrPheArgAspGlyGlnSerProAlaAlaGlyAsp 765  
 DB 1696 GACCAAGAAATGTTGCTCTATCTGTGGATCCGGATGCCAGAGTCCAGCAGCTGGAGAT 1755  
 QY 766 ValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyVal 785  
 DB 1756 GTCATCGATGGCTCTGACACAGTGTGGCTCTGCACGTTACGAATCTCTGGTGGAGGGGTG 1815  
 QY 786 TyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThr 805  
 DB 1816 TACATTTTCCACTTCGGAGTCACCCAGTCCAGGGGGCTCGGACACAGACTGCCACT 1875  
 QY 806 ValGluValGlnProAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGly 825  
 DB 1876 GTGGAAGTGCAGCCAGACCCCTAGGAAGAGTGGCTGGTGGAGCTGACCCCTGCAGGTTGGT 1935  
 QY 826 ValGlyGlnLeuThrGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeu 845  
 DB 1936 GTTGGCAGCTGCACAGACGCGGAAGGACACCCCTTGTGAGGCGAGCTGGCTGTGTGCTG 1995  
 QY 846 AsnValLeuAspSerAspIleLysValGlnLysValIleArgAlaHisSerAspLeuSerThr 865  
 DB 1996 AACGTGCTGGACTCGGATTAAGTCCAGAGATTCGGGGCCACTCGGATCTCAGACCC 2055  
 QY 866 ValIleValPheTyrValGlnSerArgProPheLysValLeuLysAlaAlaGluVal 885  
 DB 2056 GTGATGTTGTTTATGTACAGACGAGCGCGCTTTCAAGGTTCTCAAGCTGCTGAAGTG 2115  
 QY 886 AlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAspPheLeuPheLysVal 905  
 DB 2116 GCCCAAAATCTGCACATCGGCTCTCAAAAGAGAGAGCGGCTTCTTCTGCTTTTCAAGGTC 2175  
 QY 906 LeuArgValAsnThrAlaGlyCysLeuLysCysSerGlyHisGlyHisCysAspPro 925  
 DB 2176 TTGAGGGTTGATACAGAGGTTCCTCTGAAGTGTCTGGCCATGGTCACTGGCAGCCCC 2235  
 QY 926 LeuThrLysArgCysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIle 945  
 DB 2236 CTCACAAAGCGCTGCATTTGCTCTCATTATGGATGGAGAACCTTATACAGGTTATATC 2295  
 QY 946 TrpAspGlyGluSerAsnCysGluTrpSerIlePheTyrValThrValLeuAlaPheThr 965  
 DB 2296 TGGGATGGAGAGCAACTGT----- 2316  
 QY 966 LeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLys 985  
 DB 2316 ----- 2316  
 QY 986 ArgThrLysIleArgLysLysThrLysTyrThrIleLeuAspAsnMetAspGluGlnGlu 1005  
 DB 2316 ----- 2316  
 QY 1006 ArgMetGluLeuArgProLysTyrGlyIleLysHisArgSerThrGluHisAsnSerSer 1025  
 DB 2317 -----GGTATCAAGCACCGAAGCACAGACACCACTCCAGC 2352  
 QY 1026 LeuMetValSerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLys 1045  
 DB 2353 CTGATGGTATCCGAGTCTGAGTTTGCAGATGACAGGACACAAATCTTCAGCCGAGAAAAG 2412  
 QY 1046 MetGluArgGlyAsnProLys 1052  
 DB 2413 ATGGAGAGAGGGAATCCAAAG 2433

BC028869 4379 bp mRNA linear ROD 29-JUN-2004  
 Mus musculus expressed sequence AU040320, mRNA (cDNA clone  
 MGC:28660 IMAGE:4235813), complete cds.  
 BC028869  
 BC028869.1 GI:20809370  
 MGC.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 4379)  
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,  
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
 Scheetz,T.E., Brownstein,M.J., Uudin,T.B., Toshiyuki,S.,  
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
 Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,  
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
 Butlerfield,Y.S., Krzywinski,M.I., Skalka,U., Smaluk,D.E.,  
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 4379)  
 Strausberg,R.  
 Direct Submission  
 Submitted (01-MAY-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
 Contact: amg@bcm.tmc.edu  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louie, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov  
 Series: IRAK Plate: 37 Row: 1 Column: 19  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 23956263.  
 Location/Qualifiers  
 1. 4379  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="MGC:28660 IMAGE:4235813"  
 /tissue type="Kidney, normal. 5 month old male mouse."  
 /clone\_id="NCI CGAP\_Kid14"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SPORT6"



Qy 507 pSerAspGlyAlaThrAsnSerThrAlaAlaLeuLeuValAsnAlaValAspTy 527  
 Db 1664 CTCTGACGGGCTACCAACTCCACCACTGCAAGCTGACTGTGAACAAGCTGTGGACTA 1723  
 Qy 527 rProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuLeuProGlnAsnSerIleTh 547  
 Db 1724 CCCTCCCGTGGCCCAATGACAGCCCAACCAAGTATACCTGCTCAGAACTCCATCAC 1783  
 Qy 547 rLeuAsnGlyAsnGlnSerSerAspHisGlnIleValLeuLeuTyThrGluTrpSerLeuG1 567  
 Db 1784 CCTCTTTGGAAACAGAGCAGCGATGACCCAGCGCATCACAGCTATGATGCTGCCTCAG 1843  
 Qy 567 yProGlySerGluGlyHisValValMetGlnGlyValGlnThrProTyThrLeuHisLe 587  
 Db 1844 CCCGAGCAGCAAGAGGAGGTGGAGATGACGGAGTTAGAGCGCCAGCCCTGCAGCT 1903  
 Qy 587 uSerAlaMetGlnGluGlyAspTyThrPheGlnLeuLeuValThrAspSerSerArgG1 607  
 Db 1904 GTCCGCATGCAAGAGAGACTATACCTACAGCTCACAGTGAAGTGAAGTGAAGTGAAGT 1963  
 Qy 607 nGlnSerThrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAl 627  
 Db 1964 ACAGCCACCCCAAGTGAAGTGTGATGTTGTCAGCGCTGAGAACCAAGCTCTCTCAGGC 2023  
 Qy 627 aValAlaGlyProAspGlyLeuLeuIlePheProValGluSerAlaThrLeuAspGlySe 647  
 Db 2024 AGATGCAAGGCCACAGCAAGAGCTGACCTCCCGTGGACAGCAACCCCTGGACGGCAG 2083  
 Qy 647 rSerSerSerAspHisGlyIleValPheTyThrIleValArgGlyProSe 667  
 Db 2084 CAAGAGCACATGACAGAGAGTCTCTTACTTTGGAGCAGAGTCCGGGACCTGA 2143  
 Qy 667 rAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValG1 687  
 Db 2144 CGGGGTGCAGCTGGAGAAATGCAACAGCAGAGTGTCGCCACTGTGACTGGCTGCAAGTCG 2203  
 Qy 687 yThrTyHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrLe 707  
 Db 2204 GACTTATGTATTCACCTTGAATGCTCAAGATGAGAGAACTACAGAGCCAGAGCTCCGT 2263  
 Qy 707 uThrValAlaValLysLysGluLeuAsnSerProProArgAlaAlaGlyArgHi 727  
 Db 2264 TAATGTATTCTCAAGAGAAATAAACAACCCAGTACCAAGATCGCTGGGACGT 2323  
 Qy 727 sValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspG1 747  
 Db 2324 GGTGGTACCTTGGCCAGCAGCAGCTGAGCTGGATGGCTCGAGTCTCTCAGATGACAA 2383  
 Qy 747 nArgIleValSerTyThrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspVal11 767  
 Db 2384 GGGGATAGTACAGTACCTGTGAGTTCGAGATGAGAGCCAGCCAGCGGAGGAGGTGCT 2443  
 Qy 767 eAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyTh 787  
 Db 2444 GAATCACTCTGACCAACCCCTGCTCTCTCTCCAACTGTGGAGGGAGCTACAC 2503  
 Qy 787 rPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAlaThrValG1 807  
 Db 2504 GTTTCACCTGAAGTACAGATGCAAGAGGGGAGAGGAGCAGACCCGAGCAGAGTGA 2563  
 Qy 807 uValGlnProAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGlyValG1 827  
 Db 2564 AGTGAAGCTGACCCAGCAAGAAAGCAACTAGTGGAGATCATCTTGGATGTGAAGCTCAG 2623  
 Qy 827 yGlnLeuThrGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuAsnVa 847  
 Db 2624 TCAGCTGACTGAGAGGCTGAAGGGGATGCTCATCGCCAGATCTCTCTCTCTCTCTCTCT 2683  
 Qy 847 lLeuAspSerAspIleLysValGlnLysIleArgAlaHisSerAspLeuSerThrVal11 867  
 Db 2684 GCTGGATCCGACATCATTTGTGCAAAAGATTCAGCCGTACAGGAGCAGCAGCAAGAT 2743

Qy 867 eValPheTyThrValGlnSerArgProProPheLysValLeuLysAlaAlaGluValAlaAr 887  
 Db 2744 GTTCTTTTCTTTCAGAACGACCCCTCCGCCACTCTTCAAGGCCATGAGTGGCAGC 2803  
 Qy 887 gAsnLeuHisMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuAr 907  
 Db 2804 CATCTCTCAAGAGCAGCTGCAGAGCAGAGAGCTGACTTCTCATCTTCAGAGCCCTGA 2863  
 Qy 907 gValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuTh 927  
 Db 2864 AATCAGCACAGTACATGTCTGAGTGAAGTCTTCTGACCATGGCCACTGTGACTCATTCAC 2923  
 Qy 927 rLysArgCysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyThrIleTrpAs 947  
 Db 2924 CAACGCTGTGTCTGTGACCCGTTTGGATGGAATTTTCATCAAGGTGTCAGCTGAGGA 2983  
 Qy 947 pGlyLysSerAsnCysGluTrpSerIlePheTyThrValThrValLeuAlaPheThrLeu11 967  
 Db 2984 TGGAGACAGCAACTGTGAATGGAGCGTCTTACGTATCATCTTCTTGTCTATTGT 3043  
 Qy 967 eValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLysArgTh 987  
 Db 3044 TGTTCCTTGGGATCCTGTCTGACTTACAATCTGCTGTGCAAGAGCAAAA--GG 3100  
 Qy 987 rLysIleArgLysLysThrLysTyThrIleLeuAspAsnMetAspGluGlnGluArgMe 1007  
 Db 3101 AAAACCCCAAGAGAAAGCAGATACAAGATCTGATGTCACAGAT---CAGGAGAGCT 3157  
 Qy 1007 tGluLeuArgPro-----LysTyGlyIleLysHisArgSerThrGluHisAsnSerSe 1025  
 Db 3158 GGAGCTGAACCAACCTCCGAGCAGCAGCAACAGAAAGCCGCCAGCTGAGCAGGAG 3217  
 Qy 1025 rLeuMetValSerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLy 1045  
 Db 3218 CCTGATGATCTCTGAATCGGAGCTGGACAGGAC---GATGCCATCTTCATCATGSCCAGA 3274  
 Qy 1045 sMetGluArgGlyAsnProLysValSerMetAsnGlySerIleArgAsnGly 1062  
 Db 3275 CCGGAGAGAGGGCAAACTACTGTATGTCAGATGGCTCTGTGCCAAACGGG 3326  
 RESULT 5  
 BC014530 4183 bp mRNA linear PRI 29-JUN-2004  
 LOCUS BC014530  
 DEFINITION Homo sapiens polycystic kidney disease 1-like, transcript variant 1, mRNA (cdna clone MGC:11173 IMAGE:3843301), complete cds.  
 ACCESSION BC014530  
 VERSION BC014530.1 GI:45708392  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 4183)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haie, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uudin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalak, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 TITLE Human and mouse cdna sequences of more than 15,000 full-length generation and initial analysis







Db 3208 GCTGGACAGCGAT--GATGCCATCTTTACATGGCCAGACCAGAGAGAGGCAAACTCT 3264  
Qy 1052 sValSerMetAenGlySerIleArgAnGly 1062  
Db 3265 GCATGGTCAGAAATGGCTCTGTACCAACGGG 3295

## RESULT 6

LOCUS AX195570 3908 bp DNA linear PAT 28-AUG-2001  
DEFINITION Sequence 15 from Patent WO0151636.  
ACCESSION AX195570  
VERSION AX195570.1 GI:15386083

## KEYWORDS

Source Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 Lal, P., Bandman, O., Patterson, C., Walla, N.K., Nguyen, D.B., Yue, H.,  
Khan, F.A., Tang, Y.T., Baughn, M.R., Lu, D.A., Yang, J., Burford, N.,  
Au-Yang, J. and Reddy, R.

## Secreted proteins

PATENT: WO 0151636-A 15 19-JUL-2001;

Incyte Genomics, Inc. (US)

## FEATURES

Location/Qualifiers

1..3908

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

/note="Incyte ID No: 2323763CB1"

## ORIGIN

## Alignment Scores:

Pred. No.: 2.1e-120 Length: 3908  
Score: 2404.00 Matches: 522  
Percent Similarity: 60.13% Conservative: 134  
Best Local Similarity: 47.85% Mismatches: 318  
Query Match: 43.08% Indels: 117  
DB: 6 Gaps: 17

US-10-764-390-3 (1-1072) x AX195570 (1-3908)

Qy 4 ProThrGlyValLeuSerSerLeuLeuLeu-----LeuValThrIleAlaGlyCys-Al 21  
Db 296 CTGTACCTGTTTATACCTGCTTTTGTCTTACGCTTCTGTGTTGTCAACAGATGCCAG 355  
Qy 21 aArgLysGlnCysSerGluGlyArgThrTyrSerAsnAlaValIleSerProAsnLeuGl 41  
Db 356 TGAGACAGGTGCCAGCAGGGGAGACACAAATTTGGAGTTGGCTGAGATCTGGGGGAGA 415  
Qy 41 uThrThrArgIleMetArgValSerHisThrPheProValValAspCysThrAlaAlaCy 61  
Db 416 A---AATCACCCTCTGCTTCTTGAAGGAACCCCTCTCTCCAGTCATGTTGGGCTGCCTG 472  
Qy 61 sCysAspLeuSerSerCysAspLeuAlaTrpPheGluGlyArgCysTyrLeuValSe 81  
Db 473 CTGCCAGGACTCTGCTCCATGCTCTTTGGGGCTAGAGGGATGTCATTCAGGCAGA 532  
Qy 81 rCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTyrIleuth 101  
Db 533 CTGCAGCAGGCCCCAGAGCTGCCGGGCTTTAGGACACACTCTCCAAATTCATGCTGT 592  
Qy 101 rPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMetle 121  
Db 593 GTTT-----TTAAAAAAATTCACAACTGCAGATGATTG----- 626  
Qy 121 uAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArgLysAspLe 141  
Db 627 -----GCCTTTCTACCTGAGATGATGATGATGATGATGATGATGATGATGATGAT 658  
Qy 141 u\*\*\*PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTyr\*\*\*AspAspTyrAr 161

Db 659 TCTGGGCTAGTGTGAACCTGGCA----- 683  
Qy 161 gGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAlaGl 181  
Db 684 -----TCTTGAGGCGAGAGCCACCAGAGCTGCAC 715  
Qy 181 uTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGlyAs 201  
Db 716 C----- 716  
Qy 201 pSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGluSe 221  
Db 717 -AGACCTGCTGTATCTTCCAGTCAGCAGCAG----- 746  
Qy 221 rAlaSerThrProAlaProLysLeuProGluArg-----SerValLeuLeuPr 237  
Db 747 -----AGCTTAATCAGGAAGCTTCAGAGAGAGGTAGTCCAGTCAGTCAGTCACCC 799  
Qy 237 oLeuProThrThrProSerSerGlyGluValLeuGluLysGluLysAlaSerGlnLeuGl 257  
Db 800 TATAGTGACACAGCATTTCT-----AAAGTGAATGACTCCAACGAATTAGGTGGTCTGAC 853  
Qy 257 nGluGlnSerSerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProPr 277  
Db 854 TACCAGTGGCTCTGCAGAGGTCCACAGGCGATTACAATTTCCAGTCCCTTAACACACAGA 913  
Qy 277 oAlaSerLeuGluLeuSer-----SerValThrValGluLysSerPr 291  
Db 914 CTGACTGCAGAGCTGTCTGGTGGGCCAAAGAAATGATCAGTCGAACCTGAAATATCAGA 973  
Qy 291 oValLeuThrValThrProGlySer-----ThrGluHisSe 303  
Db 974 GGGCTCTGTACTACGCCCGACACTCAACAAGTAAAGATTCTGAGAAAAACCCAGATTGC 1033  
Qy 303 rIleProThrPro-----ProThrSerAlaAlaProSe 314  
Db 1034 TGTCCCCCAGCAGTGGCTCCCTCTACAGTTATGCTACCTACCCCTCCAGGCTCTTT 1093  
Qy 314 rGluSerThrProSerGluLeuProIleSerProThrThrAlaProArgThrValLysGl 334  
Db 1094 CCAGAGCACCTCAGCACCATACCCAGTT-----ATTAAGA 1129  
Qy 334 uLeuThrValSerAlaGlyAspAsnLeuIleThrLeuProAspAsnGluValGluLe 354  
Db 1130 ACTGCTGTATCTGCTGGAGAGAGTCCAGATAACCTGCCTAAGATGAAGTTCAATT 1189  
Qy 354 uLysAlaPheValAlaProAlaProProValGluThrThrTyrAsnTyrGluTrpAsnLe 374  
Db 1190 AAATGCATATGTTCTCCAAGAACCACTTAAAGGAGAAACCTACACCTACGACTGGCAGCT 1249  
Qy 374 uIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAs 394  
Db 1250 GATTACTCATCTTAGAGACTAGTGGAGAAATGGAAGGAAACATTCACAGATCTCAA 1309  
Qy 394 nLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSerGluAsnAl 414  
Db 1310 ACTATCGAAGCTCACTCCAGGCTCTGATGAATTCAAAGTGAATTGATAGAGGGTCAAAATGC 1369  
Qy 414 aPheGluGluGlyPheValAsnValThrValLysProAlaArgArgValAsnLeuProPr 434  
Db 1370 CCATGGGAAGGCTATGTGACGTGACAGTCAAGCCGAGGCCCCCGTGAAGATCGGCCCCC 1429  
Qy 434 oValAlaValValSerProGlnLeuGlnLeuThrLeuProLeuThrSerAlaLeu1 454  
Db 1430 CATTCCTATTTGTACCTCAGTTCAGGAGATCTTTTGGCCACCACTTCTACAGTCAT 1489  
Qy 454 eAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTrpGluGluLeAs 474  
Db 1490 TGAATGCGAGTCAAGACACTGATGATGATAAATCGTTAGTACCATTCATTGGGAAGAACTAA 1549  
Qy 474 nGlyProPheIleGluGluLysThrSerValAspSerProValLeuArgLeuSerAsnLe 494  
Db 1550 GGGGCTCTTAAGAGAGAAGAGAGATTCTTGAAGATACAGCCATATTAAAGTAAGTAAACT 1609



/db\_xref="GI:24559836"  
 /translation="MEKRLGVKPNPASPILSGYVWQTSKAWLSRLSLFYTFPCRSVLW  
 LSTDAESRCOQKTFQGVGLRSGGENHMLLEGTPSLQSCWAACCDPSACHVFWMLE  
 GNCIOADSRQSCRAFTRTHSNLSMLFKFPQTADDLGLFPEDDYPHLLGLGNWAS  
 WQSPRAALRPILTPKQSGSDVVPITVTHSKVNDNSNLEGLUTTS  
 GRSAEYKAITISSPLTTDLTAELSGPKNVSPQEIISGLATTSTQQVSKSEKQTIA  
 VPQVAPSYVATPFOASTAPYPIKELVVSAGESVQITLPRKNVOLNAVYVLQ  
 EPKGETVYDWQLITHPRDYSGEMGHQSLKLSKLTPLGLYEPKVTVEGONAHGEG  
 YVNVTKPEPRKNRPPIAIVSPQFOEISLPTSTTVIDGQSTDDDKIVQVHWEELKGP  
 LABEKISDEITALLKSLVPNGYTFSLTVSDSGATNTTANLTKAVDYPPVANAG  
 PNQVLTIPNSLITLKLKSLVPSKSGKVMQGVRTPIQLSAMQE  
 QDYQVLTATDTIQGOATAQVTVIOPENNPPQADAGDPDELTPVDSTLIDGSSKS  
 DDQKIISLWETQDPQDQVLENNANSSVATVGLQVTVYFTLVTKDBRNILQSSVNS  
 VIVKEINPKPTAKTNGNVITLPTSTAEIDGSKSDDKIVSYLWTRDEGSPAAGEV  
 LNHSDDHPILILSLNVECTYFHLKVDKAGESDTRTFVEKPDPRKNLVEILLDI  
 NVSQUTERLKMFIHQIGVLGLVSDIIVQIKIPTYEQSTQWVFFVQNEPPIQFKG  
 HEVAMLMKSELRKQADLIFRALEVNTVTCOLNCSHDHCDSTFKRCICDPFFWNEF  
 TKVQRDGSNCENSVLXVIATFVIVVALGILSWTVICCKRQKPKRKS KYKILD  
 ATDQESLELKPTSRAGIKQKGLILLSLMSHSEILSDDAIFTWPDREKGLLHGQNG  
 SVPNQSTPLKARSPEEIL"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,16e-120 Length: 4000  
 Score: 2404.00 Matches: 522  
 Percent Similarity: 60.13% Conservative: 134  
 Best Local Similarity: 47.85% Mismatches: 318  
 Query Match: 43.08% Indels: 117  
 DB: 9 Gaps: 17

US-10-764-390-3 (1-1072) x AV163234 (1-4000)

QY 4 ProThrGlyValLeuSerSerLeuLeuLeu-----LeuValThrIleAlaGlyCys-Al 21  
 DB 297 CCTGTACCTGTTTATCTACTTCTTTGCTTCACGCTTCTGTGTTGCTCAACAGATGCCAG 356  
 QY 21 aArgLysGlnCysSerGluGlyArgThrTyrSerAsnAlaValIleSerProAsnLeuG1 41  
 DB 357 TCAGAGCAGTGCAGCAGGAGGAGACACAAATTTGGAGTTGGCTGAGATCTGGGGGAGA 416  
 QY 41 uThrThrArgIleMetArgValSerHisThrPheProValValAspCysThrAlaAlaCy 61  
 DB 417 A---AATCACCTGTGGCTTTCTTGAAGGAACCCCTCTCTCCAGTCATGTTGGGCTGCCTG 473  
 QY 61 sCysAspLeuSerSerCysAspLeuAlaTrpThrPheGluGlyArgCysTyrLeuValSe 81  
 DB 474 CTGCCAGGACTCTGCCCTGCCATGTCCTTTGGTGGCTAGAGGAGTGTGCATTCAGGCAGA 533  
 QY 81 rCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTyrLeuTh 101  
 DB 534 CTGCAGCAGCCCGCAGAGCTGCGGGCTTTTAGGACACACTCCTCCAATTCATGCTGTGT 593  
 QY 101 rPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMetLe 121  
 DB 594 GTTTT---TTAAAAAATTCCAACTCCAGATGTTG----- 627  
 QY 121 uAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluuAspIleArgLysAspLe 141  
 DB 628 -----GGCTTTCTACCTGAAGATGATGTACCATCT 659  
 QY 141 u\*\*\*PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTyr\*\*\*AspAspTyrAr 161  
 DB 660 TCTGGGGCTAGTTGGAACCTGGCA----- 684  
 QY 161 gGluLeuGluLysAspLeuGlnProSerGlyLysGlnGluProArgGlySerAlaG1 181  
 DB 685 -----TCTTGGAGGCAGAGCCCAACCCAGAGCTGCAC 716  
 QY 181 uTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGlyAs 201  
 DB 717 C----- 717  
 QY 201 pSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGluSe 221

DB 718 -AGACCTGCTGTATCTCTCCAGTGCAGCAG-||||| 747  
 QY 221 rAlaSerThrProAlaProLysLeuProGluArg-----SerValLeuLeuPr 237  
 DB 748 -----AGCTTAATCAGGAAGCTTCGAAGAAGAGAGGTAGTCCCATGACGTAGTGTACACC 800  
 QY 237 oLeuProThrProSerSerGlyGluValLeuGluLysGluLysAlaSerGlnLeuG1 257  
 DB 801 TATAGTGACACAGCATCT-----AAAGTGAATGACTCCCAACGAATTAGGTGGTCTGAC 854  
 QY 257 nGluLnsSerSerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProPr 277  
 DB 855 TACCAGTGGCTCTGCAGAGGTCACAGAGCGATTACAAATTTCCAGTCCCTCAACACAGA 914  
 QY 277 oAlaSerLeuGluLeuSer-----SerValThrValGluLysSerPr 291  
 DB 915 CCTGACTGCAGAGCTGTCTGTGTGGCCCAAGAATGTATCAGTGCACCTGAAATATCAGA 974  
 QY 291 oValLeuThrValThrProGlySer-----ThrGluHisSe 303  
 DB 975 GGGTCTTCTACTACGCCACGACTCAACAAGTAAAGATTCTGAGAAAACCCAGATTGC 1034  
 QY 303 rIleProThrPro-----ProThrSerAlaAlaProSe 314  
 DB 1035 TGTCCCCCAGCAGTGGCTCCCTCCTACAGTATTATGCTACCCCTACCCCCAGGCTCTTT 1094  
 QY 314 rGluSerThrProSerGluLeuProIleSerProThrThrAlaProArgThrValysG1 334  
 DB 1095 CCAGAGCACCTTCAGCACCATATCCAGTT-----ATAAAGA 1130  
 QY 334 uLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAsnGluValGluLe 354  
 DB 1131 ACTGGTGTATCTCTCGAGAGAGTGTCCAGATTAACCTGCTTAAAGATGAATGAATTCAT 1190  
 QY 354 uLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTyrGluTrpAsnLe 374  
 DB 1191 AAATGCATATGTTCTCCAGAACCACCTAAAGGAGAAAACCTACACTACGACTGGCAGCT 1250  
 QY 374 uLysSerHisProThrAspTyrGlnGlyLysGlnGlyHisLysGlnThrLeuAs 394  
 DB 1251 GATTACTCATCTAGAGACTACAGTGGAGAAATGGAAGGAAACATCCAGATCCTCAA 1310  
 QY 394 nLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSerGluAsnAl 414  
 DB 1311 ACTATCGAAGTCACTCCAGGCTGTATGAATTCAAAGTGTGTAGAGGTCAAAATGC 1370  
 QY 414 aPheGlyGluGlyPheValAsnValThrValLysProAlaArgArgValAsnLeuProPr 434  
 DB 1371 CCATGGGAGAGCTATGTGAACGTGACAGTCAAGCCAGAGCCCCGTAAAGATCGGCCCC 1430  
 QY 434 oValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeu1 454  
 DB 1431 CATTCCTATTGTGCACCTCAGTTCCAGGAGATCTCTTTGGCCACACCATCTTACAGTCAT 1490  
 QY 454 eAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTrpGluGluLeAs 474  
 DB 1491 TGATGGCAGTCAAGCACTGATGATGAATAATTCGTTCAATACCATTTGGGAAGACTTAA 1550  
 QY 474 nGlyProPheLeuGluGlyLysThrSerValAspSerProValLeuLeuArgLysSerAsnLe 494  
 DB 1551 GGGGCTCTTAAGAGAAGAGAGATTTCTGAAGATACAGCCATATAAATAAGTAAACT 1610  
 QY 494 uAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSe 514  
 DB 1611 COTCCCTGGGAACACTACCTTTTCAGTGTAGTACTCTGTAGTGGAGCTACCAACTC 1670  
 QY 514 rThrThrAlaAlaLeuLeuValAsnAsnAlaValAspTyrProProValAlaAsnAlaG1 534  
 DB 1671 TACTACTGCAAACTGCAGAGTGAAACAAAGCTGTGGATTACCCCTCTGTGCCCAACGAGG 1730  
 QY 534 yProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSe 554

```
Db 1731 CCCCAACCAAGTGATCACCTGCCCCAAAACTCCATCACCTCTTTGGGAACCAAGAGCAC 1790
Qy 554 rAspAphisGlnIleValLeuThrGluTrpSerLeuGlyProGlySerGluGlyLysHi 574
Db 1791 TGATGATCATGGCATCACAGCTATGAGTGTCTACTCAGCCCAAGCAGCAAGGAAAGT 1850
Qy 574 sValValMetGlnGlyValGlnThrProTyLeuHisLeuSerAlaMetGlnGluGlyAs 594
Db 1851 GGTGAGATGAGGGTGTAGAACCAACCTTACAGCTCTCTCGATGCAAGAGGAGA 1910
Qy 594 pTyThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla***ValTh 614
Db 1911 CTACACTTACAGCTCACAGTGACTGACACATAGGACAGCAGGCCACTGCTCAAGTGAC 1970
Qy 614 rValIleValGlnProGluAenAsnArgProProValAlaValAlaGlyProAspLysG1 634
Db 1971 CGTTATTGTGCAACCTGAAACCAATAAGCCCTCTCAGGCAGATGCAGGCCAGATAAGA 2030
Qy 634 uLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerAspAspHisG1 654
Db 2031 GCTGACCCCTTCTGTGGATAGCAACCTCTGGATGGCAGCAGAGCTCAGATGATCAGAA 2090
Qy 654 yIleValPheTyHisTrpGluHisValArgGlyProSerAlaValGluMetGluAsnI1 674
Db 2091 AATTATCTCATATCTCTGGGAAAAACACAGGACCTGATGGGTGCAGCTCGAATGC 2150
Qy 674 eAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyHisPheArgLeuth 694
Db 2151 TAACAGCAGTGTGTACTGTGACTGGCTGCAAGTGGGACCTATGTGTTCACCTTGAC 2210
Qy 694 rValIleAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysG1 714
Db 2211 TGTCAAGATGAGAGAACCTGCAAGCCAGAGCTCTGTGAATGTCTGATGTCATGAAGAAGA 2270
Qy 714 uAsnAsnSerProProArgAlaArgAlaGlyGlyVargHisValLeuValLeuProAsnAs 734
Db 2271 AATAAACAACCACTATAGCAAGATAACTGGGNAATGTGTGATTACCTTACCACAG 2330
Qy 734 nSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyLeuTr 754
Db 2331 CACAGCAGCTGGATGCTCTTAAGTCTCAGATGACAAGGGAATAGTCAGCTACCTCTG 2390
Qy 754 pIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerVa 774
Db 2391 GACTCGAGATGAGGGAGCCCGCAGCAGCAGGGAGGTGTTAAATCACTCTGACCATCACCC 2450
Qy 774 lAlaLeuGlnLeuThrAsnLeuValGluGlyValTyThrPheHisLeuArgValThrAs 794
Db 2451 TATCTTTTCTTTCAAACTGGTTGAGGGAACTTACACTTTTCACTGAAAGTGACCGA 2510
Qy 794 pSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArgLy 814
Db 2511 TGCAAGGGTGAGAGTGACACAGACCGGACCACTGTGGAGGTGAAACCTGATCCAGGAA 2570
Qy 814 sSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLy 834
Db 2571 AAACAACCTGTGGAGATCATCTTGGATATCAACGTGATCAGCTCACTGAGAGGCTGAA 2630
Qy 834 sAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLysVa 854
Db 2631 GGGGATGTTCCCGCCAGATTGGGGTCTCTCTGGGGGTGCTGGATTCGACATCATTTGT 2690
Qy 854 lGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyThrValGlnSerAr 874
Db 2691 GCAAAAGATTACCGCGTACACGGACAGCAGCACAACCAATGTTATTTTGTTCAAAACGA 2750
Qy 874 gProProPheLysValLeuLysAlaAlaGluValAlaAlaArgLeuHisMetArgLeuSe 894
Db 2751 GCCTCCCAACAGATCTTCAAGGGCCATGAGTGGCAGCGATGCTCAAGAGTGAGCTCG 2810
Qy 894 rLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAlaAspThrAlaGlyCysLe 914
Db 2811 GAAGCAAAAGGAGAGACTTTTGTATATTACAGACCTTGGAAAGTCAACACTGTCACATGTCA 2870
```

```
Qy 914 uLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHi 934
Db 2871 GCTGAACCTGTTCCACCATGGCCACTGTGACTCTGTTCCACCAACGCTGTATCTGTGACCC 2930
Qy 934 sLeuTipMetGluAsnLeuIleGlnArgTyIleTrpAspGlyGluSerAsnCysGluTr 954
Db 2931 TTTTGGATGGAGAAATTTTCATCAAGGTGCACCTGAGGGATGGAGACCACTGTGAGTG 2990
Qy 954 pSerIlePheTyThrValThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheTh 974
Db 2991 GAGCGTGTATATGTATTATCTGCTACCTTTGTTCATTTGTTGCTTGTGGGAATCCTGTC 3050
Qy 974 rTrpLeuCysIleCysCysLysArgGlnLysArgThrLysIleArgLysLysThrLy 994
Db 3051 TTGGACTGTGATCTGTTGTTGTAAGAGGCAAAAA---GGAATAACCCACAGAGGAAAGCAA 3107
Qy 994 sTyThrIleLeuAspAsnMetAspGluGlnGluArgMetGluLeuArgPro-----Ly 1012
Db 3108 GTACAAGATCTGGATGCCCGAT---CAGAAAGCTGGAGCTGAAGCCAACTCCCG 3164
Qy 1012 sTyGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGluSerG1 1032
Db 3165 AGCAGGCATCAACACAGAAAGCCCTTTTGTAAAGTAGCAGCCTGATGCACTCCGAGTCAGA 3224
Qy 1032 uPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAsnProLy 1052
Db 3225 GCTGCAGCAGCGAC---GATGCCATCTTTACATGCCAGACCCAGAGAGAGGCGAATCCT 3281
Qy 1052 sValSerMetAsnGlySerIleArgAsnGly 1062
Db 3282 GCATGGTCAGAAATGGCTCTGTACCAACGGG 3312

RESULT 8
AF275679 AF275679 4001 bp mRNA linear PRI 15-JAN-2002
LOCUS Homo sapiens PP791 protein mRNA, complete cds.
DEFINITION AF275679
ACCESSION AF275679
VERSION AF275679.1 GI:10880790
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4001)
AUTHORS Qin,W.X., Wan,D.F., Zhou,X.M., Zhang,P.P., Jiang,H.Q., Huang,Y.,
Zhao,X.T. and Gu,J.R.
Novel human cDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4001)
AUTHORS Qin,W.X., Wan,D.F., Zhou,X.M., Zhang,P.P., Jiang,H.Q., Huang,Y.,
Zhao,X.T. and Gu,J.R.
Direct Submission
JOURNAL Submitted (07-JUN-2000) National Laboratory For Oncogenes & Related
Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Fu Road, Shanghai
200032, P.R. China
FEATURES
Location/Qualifiers
source 1..4001
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
208..2130
/codon_start=1
/product="pp791 protein"
/protein_id="AAG24389.1"
/db_xref="GI:10880791"
/translaton="MEKRLGVKPNPASWILSGYTWQTSAKWLRLYLFTYFCPSVLW
LSTDASSRCQQKTFQVGLRSGENHLWLEGTPLSQSCNAACCDPSACHVFWMLE
GMCTQADCSRQPSQCRAPFRSSNSMLVFLKKFTADDLGFLPDDVPHLLGLGMWNAS
WRQSPRAALRLPAVSSDDQSLIRLQKRGSPSDVWTPITVQYSKVDNSNLGLGTTTS
GSAPVHKATITSSPLTDLTALSLGSPKNSVQPEISEGLATPTSTQCVKSEKQTIA
VQPVPAPSYSTATPTPQASPOSTSAPFVVKELVVSAGESVQITLIPKNEVQLNAVLYLQ
```

EPKPGTYTVDWQIITHPRDYGEMWKGHSQILKLSUTPGLYEFKVYVBQNAHGEJ  
YVNVYVPEPRKNRPATVASFQZISLPTSTYIDGQSTDDCKTVQVHBEALGP  
LREEKISDPTAILKLSGVPNGTSLTVDSDGANTSTTANLTKVADYPPVANQ  
PNQVITLQNSHTLFGNQSTDHGTSTYEWLSPPSKKVENMGCVTPTLIQLSAQOE  
DDOYKISYLTWEGKNTGTH  
GVYQVITLQNSHTLFGNQSTDAQVTVYVQPENKNPQADAGDKELTLPVDTSTLDDGSKSS



```
Qy 614 rValIleValGlnProGluAenAenArgProProValAlaValAlaGlyProAspLysG1 634
Db 1971 CGTTATTGTCGAACCTGAAAAACAATAAGCTTCCTCAGGCAGATCGAGCCCGAGATAAGA 2030
Qy 634 uLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerAspAspHisG1 654
Db 2031 GCTGACCCCTTCCTGTGGATAGACACACACCTGGATGGCAGCAAGAGCTCAGATGATCAGAA 2090
Qy 654 yIleValPheTyrHisTrpGlu-HisValArgGlyProSerAlaValGluMetGluAenI 674
Db 2091 AATTATCTCATATCTCTGGGAAAAAACAACAGGACCTGATGGGTGCGAGTCGAGAATG 2150
Qy 674 leAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuT 694
Db 2151 CTAAACAGAGTGTGTCTACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2210
Qy 694 hrValLysAspGlnGlnGlyLeuSerSerThrThrLeuThrValAlaValLysLysG 714
Db 2211 CTGTCAAGATGAGAGGACCTGCAAGCCAGAGCTCTGTGAATGTCATTGTCAGAGAG 2270
Qy 714 luAenAenSerProProArgAlaAlaGlyArgHisValLeuValLeuProAenA 734
Db 2271 AAATAAAACAACACCTATAGCAAGATAAAGTGGGAATGTGTGATTAACCTACCCACGA 2330
Qy 734 snSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuT 754
Db 2331 GCACAGCAGAGCTGATGGCTCTAAGTCTCAGATGACAGGAATAGTACGATACCTCT 2390
Qy 754 rpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerV 774
Db 2391 GGACTCGAGATGAGGGAGCCAGCAGCAGGGAGGTGTAAATCACTCTGACCATCACC 2450
Qy 774 alAlaLeuGlnLeuThrAenLeuValGluGlyValTyrThrPheHisLeuArgValThrA 794
Db 2451 CTATCCTCTTTCTTTCAAACTGCTGTAGGAACTACACTTTTCACTGAAAGTGACCG 2510
Qy 794 spSerGlnGlyAlaSerAspThrAlaThrValGluValGlnProAspProArgL 814
Db 2511 ATGCAAGAGTGTGAGAGTGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2570
Qy 814 ysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgL 834
Db 2571 AAAACAACCTGTGTGAGATCATCTTGTGATATCAAGTCAGTCAGTCACTGAGAGGTGA 2630
Qy 834 ysAspThrLeuValArgGlnLeuAlaValLeuLeuAenValLeuAspSerAspIleLysV 854
Db 2631 AGGGGATGTTTCATCGCGCAGATTGGGTCTCTCTGGGGTCTGCGATTCGCACATCATG 2690
Qy 854 alGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSerA 874
Db 2691 TGCAAAAGATTTCAGCCGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2750
Qy 874 rgProProPheLysValLeuLysAlaAlaGluValAlaArgAenLeuHisMetArgLeus 894
Db 2751 AGCTTCCCAACAGATCTTCAAGCCATGAGTGGCAGCGATGCTCAAGAGTGAAGTGC 2810
Qy 894 erLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCysL 914
Db 2811 GGAAGCAAAAGCAGACATTTTGTATATTCAGAGCCTTGGAGTCAACACTGTCACATGTC 2870
Qy 914 euLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCysIleCysSerH 934
Db 2871 AGCTGAACTGTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2930
Qy 934 isLeuTrpMetGluAenLeuIleGlnArgTyrIleTrpAspGlyGluSerAenCysGluT 954
Db 2931 CTTTGTGATGAGAAATTTTCATAGGTGCGAGTGGAGGATGGAGACGACCACTGTGAGT 2990
Qy 954 rpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheT 974
Db 2991 GGAGCGTGTATATGTTATCATCTGCTACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3050
Qy 974 hrTrpLeuCysIleCysCysArgGlnLysArgThrLysIleArgLysLysThrL 994
```

```
Db 3051 CTTGAGACTGTGATCTGTGTGTAAGAGGCAAAA---GGAAAAACCCAAAGGAAAAAGCA 3107
Qy 994 yStyYrThrIleLeuAspAenMetAspGluGlnGlnArgMetGluLeuArgPro-----L 1012
Db 3108 AGTACAAAGATCTGTGATGCCACGGAT--CAGGAAAGCCTGGAGCTGAAGCAACCTCCC 3164
Qy 1012 yStyYrGlyIleLysHisArgSerThrGluHisAenSerSerLeuMetValSerGluSerG 1032
Db 3165 GAGCAGCATCAAAACAAAAAGCCCTTTTGTGTAAGTAGCAGCCTGATGCATCCGAGTCAG 3224
Qy 1032 luPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAenProL 1052
Db 3225 AGCTGGACAGCGAC--GATGCCATCTTTACATGGCCAGACCCGAGAGAGGAGGCAACTCC 3281
Qy 1052 ysValSerMetAenGlySerIleArgAenGly 1062
Db 3282 TGCATGTGTCAGATGGCTCTGTACCAACGGG 3313

RESULT 9
LOCUS CQ727300 2970 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 13234 from Patent WO02068579.
ACCESSION CQ727300
VERSION CQ727300.1 GI:42292890
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
AUTHORS Kits, such as nucleic acid arrays, comprising a majority of
TITLE humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 13234 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
location/Qualifiers
1..2970
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.: 2.88e-113 Length: 2970
Score: 2269.00 Matches: 473
Percent Similarity: 65.77% Conservative: 115
Best Local Similarity: 52.91% Mismatches: 244
Query Match: 40.66% Indels: 62
DB: 6 Gaps: 12

US-10-764-390-3 (1-1072) x CQ727300 (1-2970)
Qy 218 LeuAenGluSerAlaSerThrProAlaProLysLysLeuProGluArg-----Ser 233
Db 283 GTCAACAGTAGTCACAGCAGAGCTTAATCAGGAAGCTTCAGAAAGAGAGGTAGTCCCAAGTAC 342
Qy 234 ValLeuLeuProLeuProThrThrProSerSerGlyGluValLeuGluLysGluLysAla 253
Db 343 GTAGTTACACTATAGTGACACACACTTCT-----AAAGTGAATGACTCCCAACGAATTA 396
Qy 254 SerGlnLeuGlnGlnSerSerAenSerSerGlyLysGluValLeuMetProSerHis 273
Db 397 GGTGGTCTGACTACAGTGGCTCTGCAGAGTCCACAGGCGATTACAATTTCCAGTCCC 456
Qy 274 SerLeuProProAlaSerLeuGluLeuSer-----SerValThrVal 287
Db 457 CTAACCCAGACACCTGACTGCAGAGCTGTCTGGTGGGCCAAAGAAATGTATCATGTCACACT 516
Qy 288 GlnLysSerProValLeuThrValThrProGlySer----- 299
Db 517 GAAATATCAGAGGTCTTGTCTACTACGCCCAAGCACTCAACAAAGTAAAGTTCTGAGAAA 576
```

Qy	300	ThrGluHisSerIleProThrPro-----ProThrSer	310
Db	577	ACCAGATTGCTGTCGCCCCAGCCAGTCGGCTCCCTCTACAGTTATGCTACCCCTACCCCC	636
Qy	311	AlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaProArg	330
Db	637	CAGGCTCTTTTCGAGACACCTCGACACCATACCCAGTT-----	675
Qy	331	ThrValIysGluLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAsn	350
Db	676	---ATAAAGGAACCTGGTGTATCTGTGGAGAGAGTGTCCAGATAACCTTCCTAAGAT	732
Qy	351	GluValGluLeuIysAlaPheValAlaProAlaProValGluThrThrThrAsnTyr	370
Db	733	GAAGTTCAATTAATGATATGTCTCTCAAGAACCACCTAAAGAGAGAAACCTACACCTAC	792
Qy	371	GluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluIleIysGlnGlyHisLys	390
Db	793	GATGGCAGCTGATTACTCATCTCTAGAGACTACAGTGGAGAAATGGGAAGGAACATCC	852
Qy	391	GlnThrLeuAsnLeuSer-----GlnLeuSer	399
Db	853	CAGATCTCAAACTATCGAAGTGAATTTGCCTCTTTTCATCTGCGGGCAAAAGCTCACT	912
Qy	400	ValGlyLeuTyrValPheIysValThrValIserSerGluAsnAlaPheGlyGluGlyPhe	419
Db	913	CCAGGCTGTATGAATTAACAAAGTGATTTAGAGGGTCAAAATGCCATGCCGGAAGGCTAT	972
Qy	420	ValAsnValThrValIysProAlaArgValAsnLeuProProValAlaValValSer	439
Db	973	GTGAACGTGACAGTCAAGCCAGAGCCCGTAAGAATCGGCCCCCAATGTGTATTGTGTCA	1032
Qy	440	ProGlnLeuGlnLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSer	459
Db	1033	CCTCAGTTCCAGGAGATCTCTTTGCCAACCATTTCTACAGTCAATTGATGCGAGTCAAAGC	1092
Qy	460	ThrAspAspThrGluIleValSerTyrHisTrpGluGluLeuAsnGlyProPheIleGlu	479
Db	1093	ACTGATGATATAAAATCGTTTCAGTACCATTTGGGAAGAACTTAAGGGGCGCTCTAAGAGAA	1152
Qy	480	GluIlysthrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyIAsnTyr	499
Db	1153	GAGAAGATTTCTGGAAGATACAGCCCATATTAACCTAAGTAAACTCGTCCCTCGGAACCTAC	1212
Qy	500	SerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeu	519
Db	1213	ACTTTTCAGCTTGACTGTAGTAGACTCTGTATGGAGCTACCAACTCTACTCTGCAAACTG	1272
Qy	520	IleValAsnAsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIle	539
Db	1273	ACAAGTGAACAAAGCTGTGGATTAACCCCTCTGTGGCCAAACGAGGCCCAACCAAGTGAATC	1332
Qy	540	ThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIle	559
Db	1333	ACCTGCCCCAAAACCTCATCACCCTCTTTGGGAACCCAGAGCACTGAATGATCATGGCATC	1392
Qy	560	ValLeuTyrGluTrpSerLeuGlyProGlyIserGluGlyIysHisValValMetGlnGly	579
Db	1393	ACCAGCTATGAGTGGTCTACTCAGCCCCAAGCAGCAAAAGGAAAGTGTGGAGATGACGGGT	1452
Qy	580	ValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeu	599
Db	1453	GTTAGAACACCAACCTTACAGCTCTCTGCGATGCAAGAAGAGAGACTACACTTACCAGCTC	1512
Qy	600	LysValThrAspSerSerArgGlnGlnSerThrAla**ValThrValIleValGlnPro	619
Db	1513	ACAGTGACTGACACAATAGGACAGACGGCCACTGCTCAGTGACTGTATTGTGCAACCT	1572
Qy	620	GluAsnAsnArgProProValAlaValAlaGlyProAspIysGluLeuIlePheProVal	639
Db	1573	GAATAACAATTAAGCTCTCTCAGCGAGATGCGGCCAGATAAAGAGCTGACCTTCCTGTGTG	1632

[illegible]

```

Db      2710 AGGAAAGCAAGTACAGATCTCTGGATGCCACGGAT---CAGGAAAGCCTGGAGCTGAAG 2766
Qy      1011 Pro-----LysTyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetVal 1028
Db      2767 CCAACCTCCCGAGCAGGCATCAACAGAAAGGCTTTTGTCTAAGTAGCAGCTGATGCAC 2826
Qy      1029 SerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArg 1048
Db      2827 TCCGAGTCAGAGCTGGACAGCGAT--GATGCCATCTTTTATCATGTGCCGACGAGAGGAAG 2883
Qy      1049 GlyAsnProLysValSerMetAsnGlySerIleArgAsnGly 1062
Db      2884 GGCAAACTCTGATGTCAGATGGCTCTGTATCCCAACGGG 2925

RESULT 10
LOCUS   BD160555
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD160555
VERSION   BD160555.1 G1:27866313
KEYWORDS JP 2002191363-A/15398.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE   Primer for synthesizing full-length cDNA and use thereof
JOURNAL HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/15398
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, KEIICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
10,
PC C12P21/02, C12Q1/68/C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
source 1..2796
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 9.74e-109 Length: 2796
Score: 2184.50 Matches: 427
Percent Similarity: 73.14% Conservative: 93
Best Local Similarity: 60.06% Mismatches: 186
Query Match: 39.15% Indels: 5
DB: Gaps: 4

US-10-764-390-3 (1-1072) x BD160555 (1-2796)
Qy      354 LeuLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTyrGluThrAsn 373
Db      3 TTAATGCATATGTTCTCCAGAAACCCACTAAAGGAGAAACCTACACCTACGACTGGCAG 62
Qy      374 LeuLysSerHisProThrAspTyrGlnGlyGluLysGlnGlyHisLysGlnThrLeu 393
Db      63 CTGATTACTATCTCTAGAGTACAGTGGAGAAATGGAGGAAACATCCCAATCTC 122
Qy      394 AsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSerGluAsn 413

```

```

Db      123 AAATATCGAAGCTCACTCCAGGCTGTATGAATTTCAAGTGTATTGAGGGTCAAAAT 182
Qy      414 AlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgArgValAsnLeuPro 433
Db      183 GCCATGGGGAGGCTATGTGAACGTGACAGTCAAGCCAGAGCCCGCTAAGAATCGGCC 242
Qy      434 ProValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeu 453
Db      243 CCATTGCTATTGTGTACCTCAGTTTCCAGGAGATCTCTTTGGCAACACCTTCTACAGT 302
Qy      454 IleAspGlySerGlnSerThrAspAspThrGluLysValSerTyrHisTyrGluGluLys 473
Db      303 ATTATGGCAGTCAAGCACTGATGATGAATAATCGTTTCAGTACCATTTGGGAAGAATT 362
Qy      474 AsnGlyProPheIleGluGluLysThrSerValAspSerProValLeuLeuLeuSerAsn 493
Db      363 AAGGGGCTCTAAGAGAGAGAGATTTCTGAAGATACAGCATATTTAAACCTAAGTAAA 422
Qy      494 LeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsn 513
Db      423 CTCGTCCTCGGAACTACACTTTTCAGCTTGACTGTAGTACTCTGATGGAGCTACCAAC 482
Qy      514 SerThrThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProValAlaAsnAla 533
Db      483 TCTACTACTGCAACCTGACAGTGAACAAGCTGTGATTACCCCTCTGTGGCCAAACGCA 542
Qy      534 GlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSer 553
Db      543 GGCCCAACCAAGTGTATCCCTGCCCCCAAACTCCATCACCTCTTTGGGAACAGAGC 602
Qy      554 SerAspAspHisGlnIleValLeuTyrGluTyrSerLeuGlyProGlySerGluGlyLys 573
Db      603 ACTGATGATCATGGCATCACCAGCTATGAGTGGTCACTCAGCCCAAGCAGCAAGGGAAA 662
Qy      574 HisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGly 593
Db      663 GTGTGGAGATGCAAGGCTGTAGAACCAACCTTACAGCTCTCTGCGATGCAAGAAGA 722
Qy      594 AspTyrThrPheGlnLeuLysValThrAspSerArgGlnGlnSerThrAla***Val 613
Db      723 GACTACACTTACCAGCTCAGTGTGACATAGGACAGCAGCCAGCTGCTCAAGTG 782
Qy      614 ThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLys 633
Db      783 ACTGTTATTGTCAACCTGAAACAATAAAGCTCTCTCAGGCAGATGCGAGGCCAGATAA 842
Qy      634 GluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerSerSerAspHis 653
Db      843 GAGCTGACCTTCTCTGGGATAGCACCAACCTTGGATGGCAGCAGAGCTCAGATGATCAG 902
Qy      654 GlyIleValPheTyrHisTyrGluHisValArgGlyProSerAlaValGluMetGluAsn 673
Db      903 AAATATCTCATATCTCTGGGAAAAAACACAGAGGACTGTATGGGGTGACACTCGAGAT 962
Qy      674 IleAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeu 693
Db      963 GCTAACAGCAGTGTGTCTACTGTGCTGAGTGGGCTGCAAGTGGGACCTATGTGTACCTTG 1022
Qy      694 ThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysLys 713
Db      1023 ACTGTCAAGATGAGAGAACCTCTGCAAGGCCAGAGCTCTGTGAATGTCTTGTCAAGAA 1082
Qy      714 GluAsnAsnSerProProArgAlaArgAlaGlyArgHisValLeuLeuValLeuProAsn 733
Db      1083 GAAATAAACCAACCACTTATAGCCAGATACTGGGAATGTGGTGTATTACCTACCCACG 1142
Qy      734 AsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeu 753
Db      1143 AGCAGCAGAGCTGGATGGCTCTTAAGTCTTCAGATGACAGAGGAATAGTCAGCTACCTC 1202
Qy      754 TrpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSer 773

```

Db 1203 TGGACTCGAGATGAGGGGAGCCAGCAGCGGGAGGTGTTAAATCACTCTGACCATCAC 1262

Qy 774 ValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThr 793

Db 1263 CCTATCCCTTTCTTCAAACTGGTGGAGAACCTACATCTTTTCACTGAAGTGACC 1322

Qy 794 AspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArg 813

Db 1323 GATGCAAAAGGGTGAGAGTGACACAGCCGACCATGTGGAGGTGAACCTGATCCCAGG 1382

Qy 814 LysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArg 833

Db 1383 AAAACAAACCTGGTGAGATCATCTTGGATATCAACCTCAGTCACCTAATGAGAGGTG 1442

Qy 834 LysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLys 853

Db 1443 AHGGGGATGTTTCATCCGCAGATGGGGTCTCTCTGGGGTGCTCGATTCGACATCATTT 1502

Qy 854 ValGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSer 873

Db 1503 GTGCAAAAGATTACCGGTACAGCGGACGAGCACCACAAAATGGTATTTTGTTCAAAAC 1562

Qy 874 ArgProProPheLysValLeuLysAlaGluValAlaArgAsnLeuHisMetArgLeu 893

Db 1563 GAGCCTCCCCACAGATCTTCAAAGGCCATGAGGTGGCAGCGATGCTCAAGAGTGAGCTG 1622

Qy 894 SerLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCys 913

Db 1623 CGGAAGCAAAAGCGAGACTTTTGTATATTACAGGCTTGGAAAGTCAACACTGTCACATGT 1682

Qy 914 LeuLeuLysCysSerGlyHisGlyLysCysAspProLeuThrLysArgCysIleCysSer 933

Db 1683 CAGCTGAACATGTTCCGACCATCGCCACTGTGACTGTTCCACCAACGCTGATCTGTGAC 1742

Qy 934 HisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGluSerAsnCysGlu 953

Db 1743 CCTTTTGGATGGAGAAATTTTCATCAAGGTGCAGCTGAGGATGGAGACAGCAACTGTGAG 1802

Qy 954 TrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPhe 973

Db 1803 TGGAGCGTGTATATGTTATCATCTGCTACCTTTGTTCATTTGTTGCTTGGGAATCTCTG 1862

Qy 974 ThrTrpLeuCysIleCysCysLysArgGlnLysArgThrLysIleArgLysLysThr 993

Db 1863 TCTTGGACGTGATCTGTTGTTGAAGAGCAAAA---GGAAACCAAGAGGAAGCAAGC 1919

Qy 994 LysTyrThrIleLeuAspAsnMetAspGluGlnGluArgMetGluLeuArgPro----- 1011

Db 1920 AAGTACAGATCTCTGATGCCACGGAT---CAGGAAAGCTGGAGCTGAAGCCAACTCC 1976

Qy 1012 LysTyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGluSer 1031

Db 1977 CGAGCAGGATCAACAGAAAGCCCTTTTGTAAAGTAGCAGCTGATGCACTCCGAGTCA 2036

Qy 1032 GluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAsnPro 1051

Db 2037 GAGCTGGACAGCGAT---GATGCCATCTTTACATGCCACAGCAGCAGAGAGGCGCAAACTC 2093

Qy 1052 LysValSerMetAsnGlySerIleArgAsnGly 1062

Db 2094 CTGTCATGGTGCAGAAATGGCTCTGTACCAACGGG 2126

## RESULT 11

AX883831

LOCUS AX883831

DEFINITION Sequence 18736 bp DNA linear PAT 17-DEC-2003

ACCESSION AX883831

VERSION AX883831.1 GI:40038732

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Iehii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.

TITLE Primers for synthesising full-length cDNA and their use

JOURNAL Patent: EP 1074617-A 18736 07-FEB-2001;

RESEARCH Association for Biotechnology (JP)

FEATURES

source

1. 2796

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

96. 2171

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAE92130.1"

/db\_xref="GI:40038733"

/translation="MEGKHSQILKLSKLTPLGLYEFKVIVEGQAHGEGVYVNVTKPEP

RKNPAIIVSPQFEISLPTTSVIDGSQDDDKIVQIHWHEELKGLREBKISEDIT

AILKSLKLVGNVTFSTLVDSGATNSTTANLTNKAVDYPPVANAGPNQVITLUPQN

SITLFGNSQDDHGITSYWSLSPSSKGVEMQGVRTPTLQLSAMQSDYTYQITVT

DTGQATAOVTIVQENKPPQADAGPKELTLPVDSITLDGSKSSDDOKIISLYLW

EKTGGPQGVOLENANSSVATVGTGQVGTYYVFTLVKDERNLQSSSVNVIKKEINKP

PIAKTGNVITLFTSTAEIDGSKSSDDKGLIVSLWTRDEGSPAGEVLNHSDDHPIL

FLSNLVGTYTFHLKVTDAKESDTRTTEVVKPDPRKNNLVVEILINVSQTLERLK

GMFRQGLVLLGLVSDIIVQIKIQIPTYEQSTQWFFVQNNPQHQIFKGEVAAMLKSE

LRKQADLIPRALEVNTVQLNCSHDGCHDSFTKRICDPFWMENFIKVQLRDGDS

NCEWSLVVIAITFVIVVVALGILSWTVICCKROKPKPKSKKILPATDOESLELK

PTSAGIKQKGLLLSSSLMHSESELSDDDAIFTWPDREKGLKHQNGSVPNQGTPLK

ARSPREIL"

## ORIGIN

Alignment Scores:

Pred. No.: 9,74e-109 Length: 2796

Score: 2184.50 Matches: 427

Percent Similarity: 73.44% Conservative: 93

Best Local Similarity: 60.06% Mismatches: 186

Query Match: 39.15% Indels: 5

DB: 6 Gaps: 4

US-10-764-390-3 (1-1072) x AX883831 (1-2796)

Qy 354 LeuLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTyrGluTrpAsn 373

Db 3 TTAATGTCATATGTTCTCAAGAAACCACTAAAGGAGAAACCTACACCTACGACGGCAG 62

Qy 374 LeuIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeu 393

Db 63 CTGATTACTCATCTCTAGAGACTACAGTGGAGAAATGGAAAGGAAACATTCCTCCAGATCCTC 122

Qy 394 AsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSerGluAsn 413

Db 123 AAATATCGAAGCTCACTCCAGGCTGTATGAATTCAAAGTGATGTAGAGGGTCAAAAT 182

Qy 414 AlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgArgValAsnLeuPro 433

Db 183 GCCATGGGGAAGGCTATGTGAACGTGACAGTCAAGCCAGAGCCGCCGTAAGAATCGGCC 242

Qy 434 ProValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeu 453

Db 243 CCCATTGTATTGTGTACCTCAGTTCAGGAGATCTCTTTGCCAACCACTTCTTACAGTC 302

Qy 454 IleAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTrpGluGluIle 473

Db 303 ATTGATGGCAGTCAAGGCACTGATGATGATAAATCGITTCAGTACCATTTGGGAAGACTT 362

Qy 474 AsnGlyProPheIleGluGlyLysThrSerValAspSerProValLeuArgLeuSerAsn 493

Db 363 AAGGGGCTCTAAGAGAGAGAGAGATTTCTGAAGATACAGCCATATTTAAACCTAGATAA 422

Qy 494 LeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyValThrAsn 513

Db 423 CTCGTCCCTGGGAACCTACACTTTGAGTGTAGTACTGTAGTACTGTAGTGGAGACTCAAC 482



Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isegai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length

JOURNAL OF  
DIETETICS  
Nat. Genet. 36 (1), 40-45 (2004)  
14702028

2 Isogai, T., Ota, T., Hayaashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wadatsuma, M., Hozoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Iishi, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T. NEDO human cDNA sequencing project

3 (bases 1 to 2796)

**TITLE**  
**JOURNAL**  
**Direct Submission**  
**Submitted (23-AUG-2000)**

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5' - & 3' - end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES	Location/Qualifiers
source	1 2786

```

source
location/Qualifiers
1..2796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP3004051"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP3"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
96..2171
cds

```

CDS

# RIGIN

Alignment Scores:	9.74e-109	Length:	2796
red. No.:	2184.50	Matches:	427
core:	60.14%	Conservative:	93
Percent Similarity:	73.14%	Mismatches:	186
Local Similarity:	60.06%	Indels:	5
Query Match:	39.15%	Gaps:	4
B:	9		







```

Qy 937 MetGluAenLeuLeuGlnArgTyrIleTrpAspGlyGluSerAenCysGluTrpSerIle 956
Db 1693 ATGGAGAATTCATCAAGGTCAGCTGAGGATGAGACAGACAACTGTGAGTCGAGCGTG 1742
Qy 957 PheTyrValThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTripleu 976
Db 1743 TTATATGTTATCATTTGCTACCTTTGTCATTGTTGCTTGGGAATCCTGCTTGACT 1802
Qy 977 CysIleCysCysCysLysArgGlnLysArgThrLysIleArgLysLysLysLysLysThr 996
Db 1803 GTGATCTCTTGTGTGAAGAGGCAAAA---GGAAAAACCAAGAGAAAGCAAGTACAAG 1859
Qy 997 IleLeuAspAenMetAspGluGlnGluArgMetGluLeuArgPro-----LysTyrGly 1014
Db 1860 ATCTGGATGCCAGGAT---CAGAAAGCTGGAGCTGAAGCCAACTCCCGAGCAGGC 1916
Qy 1015 IleLysHisArgSerThrGluHisAenSerSerLeuMetValSerGluSerGluPheAsp 1034
Db 1917 ATCAACAGAAAGGCTTTTCTTAAGTAGCAGCCTGATGCACTCCGAGTCAGAGCTGGAC 1976
Qy 1035 SerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAenProLysValSer 1054
Db 1977 AGCGAT---GATGCCATCTTTACATGTCGACAGCCGAGAGAGGGCAAACTCCTGCATGGT 2033
Qy 1055 MetAenGlySerIleArgAsnGly 1062
Db 2034 CAGAATGGCTCTGTACCCCAACGGG 2057

RESULT 14
AB058740
LOCUS AB058740 4423 bp mRNA linear PRI 05-JUL-2002
DEFINITION Homo sapiens mRNA for KIAA1837 protein, partial cds.
ACCESSION AB058740
VERSION AB058740.1 Gi:14017890
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL DNA Res. 8 (2), 85-95 (2001)
MEDLINE 21245130
PUBMED 11347906
REFERENCE 2 (bases 1 to 4423)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2001) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdna@info.kazusa.or.jp)
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914
FEATURES
source
Location/Qualifiers
1..4423
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fh2204081"
/tissue_type="brain"
/clone_lib="pYO326GW"
/notes="vector:pBluescriptII SK plus"
gene 1..4423
/gene="KIAA1837"
CDS <28..1947
/gene="KIAA1837"
/notes="Start codon is not identified"
/codon_start=1
/product="KIAA1837 protein"
/protein_id="BAB47466.1"
/db_xref="GI:14017891"

```

```

/translation="VMEKHSQILKLSKLTPLGLYEFKVIYVEGQNAHGEQYVNTVKEP
PKRNPPIAIVSPQFSEISLPTSTVIGDSQSTDDDKIVQTHMEELKGPUREKISED
TALIKLSKLVPGNYTFLSLVDSGDTSTANLTANKVADYPPVAVANQEVITLPO
NSITLFGNQSTDDHGITSYEWLSPSKSKVEMQGVRTPTLQLSAMQEGDYTYQLTV
NDITGOQATQAVTVIYQPENKPPQADAGPKELTLVDSSTLLDGSXSSDDOKTISYL
WEKTQGDGVQLENANSSVATVGLQVGTVFTLVKDERNLOSQSSVNVIVKEEINK
PPAKITGNVITLPTSTAEIDGSKSDDKGIVSYLWTRDEGSPAAGEVNLNHSDDHPI
LFLSNLVEGYTHLKVTDKAGSDSDRTTVEVKPDKRKNLVEIILDINVSQUTERL
KGMPIROIGVLLGLVSDIIVQIKIPYESTKMFVQNEPPHQIFKGHEVAAMLKS
ELMRKADFLIFRALEVNTVQLNCSHDGHCDSFTKRCICDPEMFNFYKVOQRGD
SNCEWSVLYVIIATFVITVALGILSWTIVCCCKRQKQPKRKSRYKILDTQDSLEL
KPTSRAGRGPQCQSF"

ORIGIN
Alignment Scores: 8e-99 Length: 4423
Pred. No.: 2005.50 Matches: 388
Score: 74.92% Conservative: 84
Percent Similarity: 61.59% Mismatches: 155
Best Local Similarity: 35.94% Indels: 3
Query Match: 9 Gaps: 3
DB: 9

US-10-764-390-3 (1-1072) x AB058740 (1-4423)

Qy 383 GlyGluLeuLysGlnGly---HisLysGlnThrLeuAenLeuSerGlnLeuSerValGly 401
Db 22 GGATGAGTAATGGAAAGGAAACATTCACAGATCTCAAACCTATCGAAGCTCACCTCCAGGC 81
Qy 402 LeuTyrValPheLysValThrValSerSerGluAenAlaPheGlyGluGlyPheValAen 421
Db 82 CTGTATCAATTCAAAGTGATGTAGAGGGTCAAAATGCCCATGGGGAAGGTATGTGAAC 141
Qy 422 ValThrValLysProAlaArgValAenLeuProProValAlaValValSerProGln 441
Db 142 GTGACAGTCAAGCCAGAGCCCGTAAGAATCGGCCCCCATTTGCTATTGTGTACCTCAG 201
Qy 442 LeuGlnGluLeuThrLeuProLeuThrSerAlaLeuLeuAspGlySerGlnSerThrAsp 461
Db 202 TTCCAGAGATCTCTTTGCCAACCACTTCTACAGTCATGTATGTCAGTCAAGCACTGAT 261
Qy 462 AspThrGluLeuValSerTyrHisTyrGluGluLeuAenGlyProPheIleGluGluLys 481
Db 262 GATGATAAATCGTTCAGTACCATTTGGGAAGAACTTAAGGGGCTCTTAAGAGAGAGAGAG 321
Qy 482 ThrSerValAspSerProValLeuArgLysSerAenLeuAspProGlyAsnTyrSerPhe 501
Db 322 ATTTCTGAAGATACAGCCATATTAAATAAGTAAACTCGTCCCTCGGAACTACACTTTC 381
Qy 502 ArgLeuThrValThrAspSerAspGlyAlaThrAenSerThrThrAlaAlaLeuLeuVal 521
Db 382 AGCTTGACTGTAGTAGACTCTGATGGAGCTACCAACTCTACTTGCACAACTGCACAGTG 441
Qy 522 AsnAenAlaValAspTyrProProValAlaAenAlaGlyProAenHisThrIleThrLeu 541
Db 442 AACAAAGCTGTGGATTACCCCTCTGTGCGCAACGACGCCCCCAACCAAGTATCACCTGT 501
Qy 542 ProGlnAenSerIleThrLeuAenGlyAenGlnSerSerAspHisGlnIleValLeu 561
Db 502 CCCCAAAATCCATCCCTCTTTGGGAACACAGAGCACTGATGATCATGTCATCCAGCAGC 561
Qy 562 TyrGluTrpSerLeuGlyProGlySerGluGlyLysHisValValMetGlnGlyValGln 581
Db 562 TATGAGTGTCTACTCAGCCCAAGCAGCAAAAGGGAAGTGGTGGAGATGTCAGGGTGTAGA 621
Qy 582 ThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLysVal 601
Db 622 ACCAACTTACAGCTCTCTGCGATGCAAGAGAGAGACTACACTTACCAGCTCACAGTG 681
Qy 602 ThrAspSerSerArgGlnGlnSerThrAla***ValThrValIleValGlnProGluAen 621
Db 682 ACTGACACATAGGACAGAGCCGCTGCTCAAGTACTGTTATTGTGCAACCTGGAAC 741
Qy 622 AsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGluSer 641

```

Db	742	ATAAAGCCTCTCTCAGGACAGATGTCAGGCGCCAGATAAAGAGCTGACCCCTTCCTGCTGGATAGC	801
Qy	642	AlaThrLeuAspGlySerSerSerSerAspAspHisGlyIleValPheTyrHisTrpGlu	661
Db	802	ACAAACCTGGATGGCAGCAAGAGCTCAGATGATCAGAAAAATATCTCATATCTCTGGGAA	861
Qy	662	HisValArgGlyProSerAlaValAluMetGluAsnIleAspIysAlaIleAlaThrVal	681
Db	862	AAAAACACAGGACCTGTATGGGGTCAGCTCGAGAAATGCTAAACAGCAGTGTGTACTGTG	921
Qy	682	ThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrValIysAspGlnGlnGlyLeu	701
Db	922	ACTGGGCTGCAAGTGGGGACCTATGTGTTACCTTGACTCTCAAAAGATGAGAGCAACCTG	981
Qy	702	SerSerThrSerThrLeuThrValAlaValIysGluAsnAanSerProProArgAla	721
Db	982	CAAAAGCCAGAGCTCTGTGAATGTCATTGTCAAAAGAGAATAAACAAACACCTATATAGCC	1041
Qy	722	ArgAlaGlyGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGlySer	741
Db	1042	AAGATACTGGGAATGTGGTGATTACCTTACCACGACACAGCAGAGCTGGATGGCTCT	1101
Qy	742	ArgSerThrAspAspGlnArgIleValSerTyrIleuTrpIleArgAspGlyGlnSerPro	761
Db	1102	AAGTCCTCAGATGACAAGGGAATAGTCAGCTACCTCTGGACTCGAGATGAGGGAGGCCA	1161
Qy	762	AlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeu	781
Db	1162	GCAGCAGGGAGGGGTAAATCACTCTGACCATCACCTATTCCTTTTCTTCAAAACCTG	1221
Qy	782	ValGluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThr	801
Db	1222	GTTGAGGGAACTTACATCTTTTCACCTGAAGTGACCGATGCCAAAGGCTGAGAGTGACACA	1281
Qy	802	AspThrAlaThrValGluValGlnProAspProArgIysSerGlyLeuValGluLeuThr	821
Db	1282	GACCGGACCACTGTGGAGGTGAACCTGTATCCAGGAAACAAACCACTGGTGGAGATCATC	1341
Qy	822	LeuGlnValGlyValGlyGlnLeuThrGluGlnArgIysAspThrLeuValArgGlnLeu	841
Db	1342	TTGGATATCAACGTCAGTCAGTAACTGAGAGCGCTGAAGGGGATGTTTCATCCGCCAGATT	1401
Qy	842	AlaValLeuLeuAsnValLeuAspSerAspIleIysValGlnIysIleArgAlaHisSer	861
Db	1402	GGGGTCTCTCTGGGGTGCTGGATTCCGACATCATTTGTGCANAAGATTACGCGGTACACG	1461
Qy	862	AspLeuSerThrValIleValPheTyrValGlnSerArgProPropheIysValLeuIys	881
Db	1462	GAGCAGAGCACCAAAATGGTATTTTGTTCAAAACGAGCGCTCCCAACAGATCTTCAA	1521
Qy	882	AlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerIysGluIysAlaAspPheLeu	901
Db	1522	GGCCATGAGGTGGCAGCGCATGCTCAAGAGTGAGCTGCGGAAGCAAAAGGCAGACTTTTG	1581
Qy	902	LeuPheIysValLeuArgValAspThrAlaGlyCysLeuLeuIysCysSerGlyHisGly	921
Db	1582	ATATTAGGCCTTGGAAAGTCACACTGTCACTGTCACTTGAACCTGTTTCGACCATGGC	1641
Qy	922	HisCysAspProLeuThrIysArgCysIleCysSerHisLeuTrpMetGluAsnLeuIle	941
Db	1642	CACGTGTGACTCGTTTCCACCAACCGCTGTATCTGTGACCTTTTTCGATGGAGAATTTTCATC	1701
Qy	942	GlnArgTyrIleTrpAspGlyGluSerAsnCysGluTrpSerIlePheTyrValThrVal	961
Db	1702	AAGGTGACGCTGAGGGATGGAGACCAACTGTGAGTGGAGCGGTATATATTTATCATC	1761
Qy	962	LeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysCys	981
Db	1762	GCTACCTTTGTTCATGTGTGTGGCCTTGGGAAATCCTGTCTTGGACTGTGATCTGTGTGTGT	1821
Qy	982	LysArgGlnIysArgThrLysIleArgIysIysThrLysTyrThrIleLeuAspAsnMet	1001

Db 1822 AAGAGGCAAAA---GGAAACCACAGAGAAAGCAAGTACAAAGATCTCTGGATGCCACG 1878

Qy 1002 AspGluGlnGluArgMetGluLeuArgPro 1011  
||||| :|||:|||||:|||||

Db 1879 GAT---CAGGAAGCCTGGAGCTGAAGCCA 1905

RESULT 15

BD057820 2501 bp DNA linear PAT 27-AUG-2002

LOCUS Secreted proteins and polynucleotides encoding them.

DEFINITION

ACCESSION BD057820

VERSION BD057820.1 GI:22603426

KEYWORDS JP 2001514520-A/8.

SOURCE synthetic construct

ORGANISM synthetic construct

REFERENCE other sequences; artificial sequences.

1 (bases 1 to 2501)

AUTHORS Jacobo,K., Mccoy J.M., Lavallie,E.R., Racie,L.A., Merberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.

TITLE Secreted proteins and polynucleotides encoding them

JOURNAL GENETICS JP 2001514520-A 8 11-SEP-2001;

COMMENT PATENTS INSTITUTE INC

PN JP 2001514520-A/8

PD 11-SEP-2001

PF 09-WAR-1998 JP 1998539683

PI KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE,PI DAVID MERBERG.

PI MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC

CC Strandedness: Double;

CC Topology: Linear;

FF Key Location/Qualifiers.

FEATURES

source

1..2501

/organism="synthetic construct"

/mol\_type="genomic DNA"

/db\_xref="taxon:32630"

ORIGIN

Alignment Scores:

Pred. No.: 5,51e-89 Length: 2501

Score: 1817.50 Matches: 366

Percent Similarity: 73.23% Conservative: 80

Best Local Similarity: 60.10% Mismatches: 158

Query Match: 32.57% Indels: 7

DB: 6 Gaps: 4

US-10-764-390-3 (1-1072) x BD057820 (1-2501)

Qy 456 GlySerGlnSerThrAspAspThrGluIleValSerTyrHisTrpGluGluIleAsnGly 475  
|||||:|||||:|||||:|||||:|||||:|||||

Db 1 GGCAGTCAAAAGCACTGATGATGATGATAAAATCGTTTCAGTACCATTGGGAAGAACTTAAGGGG 60

Qy 476 ProPheIleGluGluTyrThrSerValAspSerProValLeuArgLeuSerAsnLeuAsp 495  
|||||:|||||:|||||:|||||:|||||:|||||

Db 61 CCTCTAAGAAGAAGAGAAGATTCTCGAAGATGACGACCATATTAACCTAAGTAACTCGTC 120

Qy 496 ProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThr 515  
|||||:|||||:|||||:|||||:|||||:|||||

Db 121 CCTGGGAACCTACACTTTTCAGCTTGACTGTAGTAGACTCTGATGAGACTACCAACTCTACT 180

Qy 516 ThrAlaIalaIleuIleValAsnAsnAlaValAspTyrProProValAlaAsnAlaGlyPro 535  
|||||:|||||:|||||:|||||:|||||:|||||

Db 181 ACTGCAAACTGACAGGTGAACAAAGCTGTGGATTACCCCTCTGTGGCCAAACGACGAGCCCC 240

Qy 536 AsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAsp 555  
|||||:|||||:|||||:|||||:|||||:|||||

Db 241 AACCAAGTATCATCCCTGCCCCAAAACTCCATCATCCCTCTTTGGGAACACGACGACTGAT 300

Qy 556 AspHisGlnIleValLeuTyrGluTTrpSerLeuGlyProGlySerGluGlyLysHisVal 575  
|||||:|||||:|||||:|||||:|||||:|||||

Db 301 GATCATGGCATCACCAGCTATGAGTGGTCTACTCAGCCCAAGCAGCAAGGGAAGAGTGGTG 360  
|||||:|||||:|||||:|||||:|||||:|||||

```
Qy 576 ValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluCysAspTyr 595
Db 361 GAGATGCAAGGGTGTAGAACCAACCTTACAGCTCTCTGCGATGCAAGAGGAGACTAC 420
Qy 596 ThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla**ValThrVal 615
Db 421 ACTTACCAGCTCACAGTCACTGACACAAATAGGACAGCAGCCACTGCTCAAGTGACTGT 480
Qy 616 IleValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLysGluLeu 635
Db 481 ATTGTGCAACCTGAAACAATAAGCTCTCTCAGGCAGATGCGAGGCCCAAGATAAGAGCTG 540
Qy 636 IlePheProValGlnSerAlaThrLeuAspGlySerSerSerSerSerSerSerSerSer 655
Db 541 ACCCTTCTCTGGTAGTAGCAACCTCTGGTAGGACAGAGCTCAGATGATCAGAAAAATT 600
Qy 656 ValPheTyrHisTrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAsp 675
Db 601 ATCTCATATCTCTGGGAAAAACACAGAGGACTGTATGGGTGCAGCTCGAGNATGCTAAC 660
Qy 676 LysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrVal 695
Db 661 AGCAGTGTGTACTGTACTGGCTGCAAGTGGGACCTATGTGTTCACCTTGACTGTC 720
Qy 696 LysAspGlnGlnGlyLeuSerSerThrThrThrValAlaValAlaValLysLysGluAsn 715
Db 721 AAAGATGAGAGGAACCTCAAGCCAGAGCTCTGTGAATGTCTTGTCAAAAGAGAAATA- 779
Qy 716 AsnSerProProArgAlaArgAlaGlyArgHisValLeuValLeuProAsnAsnSer 735
Db 780 AACAAACCACTATAGCCCAAGATACTGGGAATGTGGTATTACCTACCACAGGACACA 839
Qy 736 IleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuThrIle 755
Db 840 GCAGAGCTGGATGGCTCTAAGTCTCTCAGATGACAAAGGGAATAGTCAGCTACTCTGGACT 899
Qy 756 ArgAspGlyGlnSerProAlaalaGlyAspValIleAspGlySerAspHisSerValAla 775
Db 900 CGAGTAGGGGGAGCCCGACGAGCAGGAGGGTGTAAATCACTCTGACCACTCAACCCCTATC 959
Qy 776 LeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrAspSer 795
Db 960 CTTTTTCTTTCAAACTCTGGTGGAGGAACCTACACTTTTTCACCTGAAAGTGACCGATGCA 1019
Qy 796 GlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArgLysSer 815
Db 1020 AAGGCTGAGAGTGACACAGACCGGACCCTGTGGAGGTGAAACCTGTATCCAGGAAAAAC 1079
Qy 816 GlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAsp 835
Db 1080 AACCTGGTGGAGATCATCTTGGATATCAACCGTCACTCAACTGAGAGGCTGAAGGGG 1139
Qy 836 ThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLysValGln 855
Db 1140 ATGTTTCATCCGCGAGATTGGGGTCTCTCTGGGGGTGCTGGATTCCGACATCATTTGTGCAA 1199
Qy 856 LysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSerArgPro 875
Db 1200 AAGATTACGCGGTACACGAGCAGACACCAAAATGGTATTTTGTTCAAAACGAGCCT 1259
Qy 876 PropheLysValLeuLysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerLys 895
Db 1260 CCCCACCAAGATTTCAAAAGGCCCATGAGGTGGCAGCGATGCTCAAGAGTGAGCTGCGAAG 1319
Qy 896 GluLysAlaAspPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeu 915
Db 1320 CAAAAGGCAGACTTTTGTATTTACAGCCTTGGAAAGTCAACACTGTCTACATGTCAAGCTG 1379
Qy 916 LysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeu 935
Db 1380 AACTGTTCCGACCATGGCTACTGTGACTCGTTCACCAACCGCTGTATCTGTGACCCCTTT 1439
Qy 936 TrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGluSerAsnCysGluTrpSer 955
```

```
Db 1440 TGGATGAGAGATTTCATCAAGTCCAGCTGAGGATGGAGACAGCAACTGTGATGGAGC 1499
Qy 956 IlePheTyrValThrValLeuAlaPheThrIleValLeuThrGlyGlyPheThrTrp 975
Db 1500 GTGTTATATGTTATCAITGTCTACCTTTGTCTTGTTCATTGTGTGGTGGGAATCCTGTCTGG 1559
Qy 976 LeuCysIleCysCysCysValLeuArgGlnLysArgThrLysIleArgLysLysThrLysTyr 995
Db 1560 ACTGTGATCTGTTGTTGTAAGAGCAAAAA---GGAAAAACCAAGAGGAAAGCAAGTAC 1616
Qy 996 ThrIleLeuAspAsnMetAspGluGlnGluArgMetGluLeuArgPro-----LysTyr 1013
Db 1617 AAGAT-CTGGATGCCACGGAT--CAGGAAAGCTTGGAGCTGAAGCCCAACCTCCCGACA 1672
Qy 1014 GlyIleLysHisArgSerThrGluHisAsnSerSerSerLeuMetValSerGluSerGluPhe 1033
Db 1673 GGCATCAACACAGAAAGGCCCTTTTGTCTAAGTAGCAGCCTGTATGCACCTCCGAGTCAAGCTG 1732
Qy 1034 AspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAsnProLysVal 1053
Db 1733 GACAGCGAT---GATGCCATCTTTTACATGGCCAGACCGAGAGAGGGCAAACTCCTCTGCAT 1789
Qy 1054 SerMetAsnGlySerIleArgAsnGly 1062
Db 1790 GGTCAAGATGGCTCTGTACCAACGGG 1816
```

Search completed: October 12, 2005, 13:41:44  
Job time : 10126 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:00:24 ; Search time 125 Seconds

(without alignments)  
3316.857 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSLLLVTTIAGC.....VSMNGSIRNGASFYSCKSDR 1072

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_16Dec04:\*

- 1: Geneseq\_1980s:\*
- 2: Geneseq\_1980s:\*
- 3: Geneseq\_2000s:\*
- 4: Geneseq\_2001s:\*
- 5: Geneseq\_2002s:\*
- 6: Geneseq\_2003as:\*
- 7: Geneseq\_2003bs:\*
- 8: Geneseq\_2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5576	99.9	1072	7 ADJ69816	Adj69816 Human hea
2	5576	99.9	1072	8 ADR00600	Adr00600 Human 254
3	5574	99.9	1072	8 ADR00601	Adr00601 Human 254
4	5574	99.9	1072	8 ADR00597	Adr00597 Human 254
5	5574	99.9	1072	8 ADR00598	Adr00598 Human 254
6	5574	99.9	1072	8 ADR00848	Adr00848 Human 254
7	5574	99.9	1072	8 ADR00594	Adr00594 Human 254
8	5574	99.9	1072	8 ADR00592	Adr00592 Human 254
9	5485	98.3	1053	8 ADR00862	Adr00862 Human 254
10	5485	98.3	1053	8 ADR00863	Adr00863 Human 254
11	5485	98.3	1063	8 ADR00861	Adr00861 Human 254
12	5485	98.3	1063	8 ADR00596	Adr00596 Human 254
13	5485	98.3	1063	8 ADR00599	Adr00599 Human 254
14	4938	88.5	978	7 ADG14994	Adg14994 Human SEC
15	2426.5	43.5	1048	8 ADM36227	Adm36227 Murine SS
16	2401	43.0	1049	4 AAE05768	Aae05768 Human sec
17	2401	43.0	1049	7 ADI21080	Adi21080 Novel hum
18	2086.5	37.4	691	4 AAB95784	Aab95784 Human pro
19	1431.5	26.0	1069	4 ABB63364	Abb63364 Drosophil
20	1440.5	25.8	685	4 AAB94231	Aab94231 Human pro
21	1375	24.6	711	4 ABG08994	Abg08994 Novel hum
22	1296.5	23.2	640	7 ADP42454	Adp42454 Human pp7
23	823	14.7	272	6 ABU70371	Abu70371 Human adi
24	815.5	14.6	346	4 AAM00902	Aam00902 Human bon
25	487	8.7	162	4 AAO09933	Aao09933 Human pol

#### ALIGNMENTS

##### RESULT 1

ADJ69816

ID ADJ69816 standard; protein; 1072 AA.

XX AC ADJ69816;

DT 06-MAY-2004 (first entry)

DE Human heat mitochondrial protein as a therapeutic target SeqID1622.

XX KW mitochondrial; human; screening assay; diabetes mellitus;

XX KW Huntington's disease; osteoarthritis;

XX KW Leber's hereditary optic neuropathy; LHON;

XX KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;

XX KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;

XX KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;

XX KW osteopathic; ophthalmological; cytostatic.

XX OS Homo sapiens.

XX XX WO2003087768-A2.

PN PD 23-OCT-2003.

XX PF 04-APR-2003; 2003WO-US010870.

XX PR 12-APR-2002; 2002US-0372843P.

XX PR 17-JUN-2002; 2002US-0389987P.

XX PR 20-SEP-2002; 2002US-0412418P.

XX PA (MITO-) MITOKOR.

XX PA (BUCK-) BUCK INST AGE RES.

XX PI Ghosh SS, Pahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX PI Warnock DE;

XX DR WPI; 2003-845369/78.

XX XX Identifying a mitochondrial target for drug screening assays and for

XX PT treating diseases associated with altered mitochondrial function,

XX PT comprises detecting a modified polypeptide in a sample and correlating

XX PT with the disease.

XX XX Claim 1; SEQ ID NO 1622; 180pp; English.

XX PS This invention relates to novel mitochondrial targets that can be used

XX PS for therapeutic intervention in treating a disease associated with

XX CC altered mitochondrial function. Specifically, it refers to a method for

XX CC altered mitochondrial function.

XX CC

identifying proteins of the human heart mitochondrial proteome that are useful for drug screening assays, as well as therapeutic targets. The present invention describes a method for identifying such proteins that can be used in the treatment of various diseases associated with altered mitochondrial function including diabetes mellitus, Huntington's disease, osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy ragged red fibre syndrome (MERRF) or cancer. Accordingly, these compositions have neuroprotective, neurotropic, antidiabetic, anticonvulsant, antirheumatic, osteopathic, ophthalmological and cytostatic activities. This polypeptide sequence is a human heart mitochondrial protein of the invention.

XX Sequence 1072 AA;

Query Match 99.9%; Score 5576; DB 7; Length 1072;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAPPTGVLSLLILLITIAACARKQCEGRYSNAVISPNLETRIMRVSHTFPVVDTAA 60  
DB 1 MAPPTGVLSLLILLITIAACARKQCEGRYSNAVISPNLETRIMRVSHTFPVVDTAA 60  
QY 61 CDDLSDDLAWFEGRCYLVSCPHKENCEPKMGPIRSYLVFLVPVORPAQLLDYDGM 120  
DB 61 CDDLSDDLAWFEGRCYLVSCPHKENCEPKMGPIRSYLVFLVPVORPAQLLDYDGM 120  
QY 121 LNRGSPGSGWDPEDIRKDLFLGKDWGLEMSYDXYREKDLQPSGKQPRGSA 180  
DB 121 LNRGSPGSGWDPEDIRKDLFLGKDWGLEMSYDXYREKDLQPSGKQPRGSA 180  
QY 181 EYTDWGLLPGSEGANSSVGSVPAPVPAETQDQPELHYLNESASTAPKLPERSVLLPPT 240  
DB 181 EYTDWGLLPGSEGANSSVGSVPAPVPAETQDQPELHYLNESASTAPKLPERSVLLPPT 240  
QY 241 TPSSGEVLEKAKASQLQGSNNSSKEVLMPSHSLPPASLELSSVTVTKSPVLTVTPGST 300  
DB 241 TPSSGEVLEKAKASQLQGSNNSSKEVLMPSHSLPPASLELSSVTVTKSPVLTVTPGST 300  
QY 301 EHSIPTPTSAAPSESTSELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360  
DB 301 EHSIPTPTSAAPSESTSELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360  
QY 361 APPVETTYNENLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNARFGEFV 420  
DB 361 APPVETTYNENLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNARFGEFV 420  
QY 421 NVTVKPARRVNLPPVAVVSPQLETLPLTSALIDGQSTDDTEIVSVHWEELNPFLEE 480  
DB 421 NVTVKPARRVNLPPVAVVSPQLETLPLTSALIDGQSTDDTEIVSVHWEELNPFLEE 480  
QY 481 KTSVDSPLVRLSNLDPGNSYFRLTVTSDGATNSTTAALIVNNAVDPYPPVANAGPNHIT 540  
DB 481 KTSVDSPLVRLSNLDPGNSYFRLTVTSDGATNSTTAALIVNNAVDPYPPVANAGPNHIT 540  
QY 541 LPQNSITLNGOSSDHDQIVLYEWSLPGSEKGVVMQVQPPYHLHSAMQEGDYTFOLK 600  
DB 541 LPQNSITLNGOSSDHDQIVLYEWSLPGSEKGVVMQVQPPYHLHSAMQEGDYTFOLK 600  
QY 601 VTDSRQOSTAXVTIVOPENNRPPVAVAGDPKELIPVESATLDGSSDDHGVFVHM 660  
DB 601 VTDSRQOSTAXVTIVOPENNRPPVAVAGDPKELIPVESATLDGSSDDHGVFVHM 660  
QY 661 EHVRFPSAVEMENIDKATATVGLQVTHFLRTVKDQOGLSSTSTLTAVVKENNSPPR 720  
DB 661 EHVRFPSAVEMENIDKATATVGLQVTHFLRTVKDQOGLSSTSTLTAVVKENNSPPR 720  
QY 721 ARAGGRHVLVLPNNISITLDGSRSTDDQRTVSVLWIRDQGSAAAGVDVIGSDHVALQLTN 780  
DB 721 ARAGGRHVLVLPNNISITLDGSRSTDDQRTVSVLWIRDQGSAAAGVDVIGSDHVALQLTN 780  
QY 781 LVEGVYTHLRVTDGASDATTATVEQDPKSGSLVELTLQVGVGLTEQRKDTLVRQ 840

DB 781 LVEGVYTHLRVTDGASDATTATVEQDPKSGSLVELTLQVGVGLTEQRKDTLVRQ 840  
QY 841 LAVLLNVLDSIDIKVQKTRAHSDLSSTVIVFYVQSPFPFKVLAEEVARNLHRLSEKADP 900  
DB 841 LAVLLNVLDSIDIKVQKTRAHSDLSSTVIVFYVQSPFPFKVLAEEVARNLHRLSEKADP 900  
QY 901 LLFKVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYTIWDGESNCEWSIFYVT 960  
DB 901 LLFKVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYTIWDGESNCEWSIFYVT 960  
QY 961 VLAFTLIVLTGFTWLTCICCCCKRQKRTKIRKTKYTTILDNMDEQERMBELRPYGIKHRST 1020  
DB 961 VLAFTLIVLTGFTWLTCICCCCKRQKRTKIRKTKYTTILDNMDEQERMBELRPYGIKHRST 1020  
QY 1021 EHNSSLVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKDR 1072  
DB 1021 EHNSSLVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKDR 1072  
RESULT 2  
ADR00600  
ID ADR00600 standard; protein; 1072 AA.  
XX AC ADR00600;  
XX AC ADR00600;  
DT 04-NOV-2004 (first entry)  
XX Human 254P1D6B v.5 protein SEQ ID NO:11.  
DE 254P1D6B; small interfering RNA; siRNA; immune response;  
KW 254P1D6B-related protein; cytotostatic; gene therapy; cancer; human;  
KW 254P1D6B v.5; chromosome 6.  
XX Homo sapiens.  
XX WO2004067716-A2.  
XX 12-AUG-2004.  
XX 23-JAN-2004; 2004WO-US001965.  
PR 24-JAN-2003; 2003US-0442526P.  
XX (AGEN-) AGENSYS INC.  
PI Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
PI Perez-Villar JJ, Paris M;  
XX WPI; 2004-580991/56.  
XX New 254P1D6B siRNA composition comprising a double stranded siRNA that corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B protein or corresponds to a subsequence of the ORF, useful for detecting and treating cancer.  
PS Example 1; SEQ ID NO 11; 345pp; English.  
XX The present invention describes a 254P1D6B small interfering RNA (siRNA) composition that comprises a double stranded siRNA that corresponds to the nucleic acid open reading frame (ORF) sequence which encodes the 254P1D6B protein, or corresponds to a subsequence of the ORF, where the double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous nucleotides in length. Also described: (1) a composition that comprises, consists essentially of, or consists of a peptide of eight, nine, ten, or eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID NO.3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in any of the 42 lists of peptides, given in the specification, or a protein that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or identical to an entire amino acid sequence of PI; (2) a polynucleotide that encodes the protein; (3) a composition comprising a polynucleotide that is a full complement of the polynucleotide described above; (4) generating a mammalian immune response directed to the protein of PI; (5) detecting, in a sample, the presence of a 254P1D6B-related protein or a

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:24:05 ; Search time 73 Seconds

(without alignments)  
5679.550 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSLLLVTTIAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 854512

Minimum DB seq length: 0

Maximum DB seq length: 35

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 16Dec04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	158	2.8	29	8	ADR00857 Human 254
2	152	2.7	29	8	ADR00851 Human 254
3	101	1.8	19	8	ADR00850 Human 254
4	99	1.8	19	8	ADR00856 Human 254
5	92	1.6	16	8	ADR00632 Human 254
6	90	1.6	17	8	ADR00855 Human 254
7	84	1.5	17	8	ADR00849 Human 254
8	83	1.5	19	4	ABG22066 Novel hum
9	79	1.4	24	8	ADR00854 Human 254
10	63	1.1	9	8	ADR00727 Human 254
11	61.5	1.1	34	4	AB85615 Monoclonal
12	61.5	1.1	34	5	AB85625 Cerebrosp
13	60	1.1	12	8	ADR00630 Human 254
14	59	1.1	9	8	ADR00726 Human 254
15	59	1.1	9	8	ADR00838 Human 254
16	57	1.0	9	8	ADR00641 Human 254
17	57	1.0	31	3	AB225609 Peptide e
18	57	1.0	31	6	ADA27081 Human nov
19	57	1.0	31	8	ABE86613 Integrin
20	57	1.0	34	8	ABO57279 Human gen
21	56	1.0	9	8	ADR00814 Human 254
22	56	1.0	9	8	ADR00728 Human 254
23	56	1.0	9	8	ADR00829 Human 254
24	55	1.0	9	8	ADR00639 Human 254
25	55	1.0	9	8	ADR00675 Human 254

26	55	1.0	9	8	ADR00828 Human 254
27	55	1.0	11	8	ADR00631 Human 254
28	55	1.0	33	3	AAB25603 Peptide e
29	55	1.0	33	6	ADA27075 Human nov
30	55	1.0	33	8	ADE86607 Integrin
31	54.5	1.0	23	7	ADC35432 FITC label
32	54.5	1.0	23	7	ADC35439 FAM label
33	54	1.0	9	8	ADR00694 Human 254
34	54	1.0	9	8	ADR00826 Human 254
35	54	1.0	9	8	ADR00725 Human 254
36	54	1.0	9	8	ADR00747 Human 254
37	54	1.0	9	8	ADR00810 Human 254
38	54	1.0	29	1	AAP81554 Human ins
39	54	1.0	29	8	ADI80581 Conjugate
40	54	1.0	34	3	AAB25607 Peptide e
41	54	1.0	34	6	ADA27079 Human nov
42	54	1.0	34	8	ADE86611 Integrin
43	53.5	1.0	24	5	ABB08283 Human Gal
44	53.5	1.0	24	6	ABU07534 Human N-a
45	53.5	1.0	24	7	ADE25869 Human muc

## ALIGNMENTS

## RESULT 1

ADR00857  
ID ADR00857 standard; peptide; 29 AA.

XX ADR00857;

XX 04-NOV-2004 (first entry)

XX DE Human 254PID6B v.5 peptide SEQ ID NO:268.

XX KW 254PID6B; small interfering RNA; siRNA; immune response;

KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;

KW 254PID6B v.5; chromosome 6.

XX OS Homo sapiens.

XX PN W02004067716-A2.

XX PD 12-AUG-2004.

XX PF 23-JAN-2004; 2004WO-US001965.

XX PR 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;

XX Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.

XX New 254PID6B siRNA composition comprising a double stranded siRNA that corresponds to the nucleic acid ORF sequence which encodes the 254PID6B protein or corresponds to a subsequence of the ORF, useful for detecting and treating cancer.

XX Example 13; SEQ ID NO 268; 345pp; English.

XX The present invention describes a 254PID6B small interfering RNA (siRNA) composition that comprises a double stranded siRNA that corresponds to the nucleic acid open reading frame (ORF) sequence which encodes the 254PID6B protein, or corresponds to a subsequence of the ORF, where the double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous nucleotides in length. Also described: (1) a composition that comprises, consists essentially of, or consists of a peptide of eight, nine, ten, or eleven contiguous amino acids of a protein of figure 2 (p1, see SEQ ID NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in any of the 42 lists of peptides, given in the specification, or a protein

CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents a  
 CC human 254P1D6B v.5 peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 CC  
 CC Sequence 29 AA;

Query Match 2.8%; Score 159; DB 8; Length 29;  
 Best Local Similarity 96.6%; Pred. No. 0.00021;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 128 GIWGSPEDIRKDLFLGKDWGLEMSEY 156  
 |||||  
 DB 1 GIWGSPEDIRKDLFLGKDWGLEMSEY 29

RESULT 2  
 ADR00851  
 ID ADR00851 standard; peptide; 29 AA.

XX ADR00851;

XX 04-NOV-2004 (first entry)

XX Human 254P1D6B v.2 peptide SEQ ID NO:262.

XX 254P1D6B; small interfering RNA; siRNA; immune response;  
 KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254P1D6B v.2; chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.

XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX Example 13; SEQ ID NO 262; 345pp; English.

XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the

CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents a  
 CC human 254P1D6B v.2 peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 CC  
 CC Sequence 29 AA;

Query Match 2.7%; Score 152; DB 8; Length 29;  
 Best Local Similarity 96.6%; Pred. No. 0.00059;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 143 FLGKDWGLEMSEYKDDYRELEKDLLOPS 171  
 |||||  
 DB 1 FLGKDWGLEMSEYADYRELEKDLLOPS 29

RESULT 3  
 ADR00850

ID ADR00850 standard; peptide; 19 AA.

XX ADR00850;

XX 04-NOV-2004 (first entry)

XX Human 254P1D6B v.2 peptide SEQ ID NO:261.

XX 254P1D6B; small interfering RNA; siRNA; immune response;  
 KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254P1D6B v.2; chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.

XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

PS Example 13; SEQ ID NO 261; 345pp; English.

XX The present invention describes a 254P1D6B small interfering RNA (siRNA) that corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B protein or corresponds to a subsequence of the ORF, useful for detecting and treating cancer.

CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous nucleotides in length. Also described: (1) a composition that comprises, consists essentially of, or consists of a peptide of eight, nine, ten, or eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in any of the 42 lists of peptides, given in the specification, or a protein that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or identical to an entire amino acid sequence of P1; (2) a polynucleotide that encodes the protein; (3) a composition comprising a polynucleotide that is a full complement of the polynucleotide described above; (4) generating a mammalian immune response directed to the protein of P1; (5) detecting, in a sample, the presence of a 254P1D6B-related protein or a 254P1D6B-related polynucleotide; (6) a composition that modulates the status of a cell that expresses a protein of P1; (7) a pharmaceutical composition that comprises the composition described above in a human unit dose form; (8) an antibody or its fragment, which is monoclonal; (9) a non-human transgenic animal that produces the antibody; (10) a hybridoma that produces the antibody; (11) delivering a cytotoxic agent or a diagnostic agent to a cell that expresses the protein of P1; and (12) inhibiting growth, reproduction or survival of cancer cells that express the protein of P1. 254P1D6B has cytostatic activity, and can be used in gene therapy. The compositions, molecules and methods are useful for treating and detecting cancer. The present sequence represents a human 254P1D6B v.2 peptide, which is used in the exemplification of the present invention. The human 254P1D6B gene is located on chromosome 6p22.

XX Sequence 19 AA;

Qy 148 WGLEEMSEYDDYRELEKD 166

Db 1 WGLEEMSEYDDYRELEKD 19

RESULT 4

ADRO0856

ID ADR00856 standard; peptide; 19 AA.

XX AC ADR00856;

XX 04-NOV-2004 (first entry)

XX Human 254P1D6B v.5 peptide SEQ ID NO:267.

XX 254P1D6B; small interfering RNA; siRNA; immune response;

XX 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;

XX 254P1D6B v.5; chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;

XX Perez-Villar JJ, Paris M;

XX WPI; 2004-580991/56.

XX New 254P1D6B siRNA composition comprising a double stranded siRNA that corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B protein or corresponds to a subsequence of the ORF, useful for detecting and treating cancer.

XX Example 13; SEQ ID NO 267; 345pp; English.

XX The present invention describes a 254P1D6B small interfering RNA (siRNA) that corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B protein or corresponds to a subsequence of the ORF, where the double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous nucleotides in length. Also described: (1) a composition that comprises, consists essentially of, or consists of a peptide of eight, nine, ten, or eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in any of the 42 lists of peptides, given in the specification, or a protein that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or identical to an entire amino acid sequence of P1; (2) a polynucleotide that encodes the protein; (3) a composition comprising a polynucleotide that is a full complement of the polynucleotide described above; (4) generating a mammalian immune response directed to the protein of P1; (5) detecting, in a sample, the presence of a 254P1D6B-related protein or a 254P1D6B-related polynucleotide; (6) a composition that modulates the status of a cell that expresses a protein of P1; (7) a pharmaceutical composition that comprises the composition described above in a human unit dose form; (8) an antibody or its fragment, which is monoclonal; (9) a non-human transgenic animal that produces the antibody; (10) a hybridoma that produces the antibody; (11) delivering a cytotoxic agent or a diagnostic agent to a cell that expresses the protein of P1; and (12) inhibiting growth, reproduction or survival of cancer cells that express the protein of P1. 254P1D6B has cytostatic activity, and can be used in gene therapy. The compositions, molecules and methods are useful for treating and detecting cancer. The present sequence represents a human 254P1D6B v.5 peptide, which is used in the exemplification of the present invention. The human 254P1D6B gene is located on chromosome 6p22.

XX Sequence 19 AA;

Qy 133 SPEDIRKDLXFLGKDWGLE 151

Db 1 SPEDIRKDLXFLGKDWGLE 19

RESULT 5

ADRO0632

ID ADR00632 standard; peptide; 16 AA.

XX AC ADR00632;

XX 04-NOV-2004 (first entry)

XX Human 254P1D6B peptide SEQ ID NO:43.

XX 254P1D6B; small interfering RNA; siRNA; immune response;

XX 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;

XX chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX Query Match 1.8%; Score 99; DB 8; Length 19;

XX Best Local Similarity 94.7%; Pred. No. 2.3;

XX Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

PA (AGEN-) AGENSYS INC.  
 XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
 PI Perez-Villar JJ, Paris M;  
 XX WPI; 2004-580991/56.  
 DR  
 XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.  
 PS  
 XX Example 9; SEQ ID NO 43; 345pp; English.  
 XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254P1D6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 XX Sequence 16 AA;  
 SQ  
 Query Match 1.6%; Score 92; DB 8; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 522 NNAVDPVPVANAGPNH 537  
 Db 1 NNAVDPVPVANAGPNH 16  
 RESULT 6  
 ADR00855  
 ID ADR00855 standard; peptide; 17 AA.  
 XX ADR00855;  
 AC  
 XX 04-NOV-2004 (first entry)  
 DT  
 XX Human 254P1D6B v.5 peptide SEQ ID NO:266.  
 DE  
 XX 254P1D6B; small interfering RNA; siRNA; immune response;  
 KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254P1D6B v.5; chromosome 6.  
 XX Homo sapiens.  
 OS  
 XX WO2004067716-A2.  
 PN  
 XX

PD 12-AUG-2004.  
 XX 23-JAN-2004; 2004WO-US001965.  
 XX 24-JAN-2003; 2003US-0442526P.  
 PR (AGEN-) AGENSYS INC.  
 XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
 PI Perez-Villar JJ, Paris M;  
 XX WPI; 2004-580991/56.  
 DR  
 XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.  
 PS  
 XX Example 13; SEQ ID NO 266; 345pp; English.  
 XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents a  
 CC human 254P1D6B v.5 peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 XX Sequence 17 AA;  
 SQ  
 Query Match 1.6%; Score 90; DB 8; Length 17;  
 Best Local Similarity 94.1%; Pred. No. 8.7;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 134 PEDIRKDLXFLGKDWGL 150  
 Db 1 PEDIRKDLTFLGKDWGL 17  
 RESULT 7  
 ADR00849  
 ID ADR00849 standard; peptide; 17 AA.  
 XX ADR00849;  
 AC  
 XX 04-NOV-2004 (first entry)  
 DT  
 XX Human 254P1D6B v.2 peptide SEQ ID NO:260.  
 DE  
 XX 254P1D6B; small interfering RNA; siRNA; immune response;  
 KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW



KW 254P1D6B v.2; chromosome 6.  
 XX Homo sapiens.  
 OS  
 XX WO2004067716-A2.  
 FN  
 XX 12-AUG-2004.  
 PD  
 XX  
 XX 23-JAN-2004; 2004WO-US001965.  
 PF  
 XX 24-JAN-2003; 2003US-0442526P.  
 PR  
 XX (AGEN-) AGENSYS INC.  
 PA  
 XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
 FI Perez-Villar JJ, Paris M;  
 PI  
 XX WPI; 2004-580991/56.  
 DR  
 XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 FT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.  
 XX  
 XX Example 13; SEQ ID NO 260; 345pp; English.  
 PS  
 XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents a  
 CC human 254P1D6B v.2 peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 XX  
 SQ Sequence 17 AA;  
 Query Match 1.5%; Score 84; DB 8; Length 17;  
 Best Local Similarity 94.1%; Pred. No. 24;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 149 GLEEMSEYDDYRELEK 165  
 Db 1 GLEEMSEYDDYRELEK 17  
 RESULT 8  
 ABG22066  
 ID ABG22066 standard; protein; 19 AA.  
 XX  
 AC ABG22066;  
 XX  
 18-FEB-2002 (first entry)  
 Novel human diagnostic protein #22057.  
 Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.  
 Homo sapiens.  
 WO200175067-A2.  
 11-OCT-2001.  
 30-MAR-2001; 2001WO-US008631.  
 31-MAR-2000; 2000US-00540217.  
 23-AUG-2000; 2000US-00649167.  
 (HYSE-) HYSEQ INC.  
 Drmanac RT, Liu C, Tang YT;  
 WPI; 2001-639362/73.  
 N-PSDB; AAS86253.  
 New isolated polynucleotide and encoded polypeptides, useful in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits and to assess  
 biodiversity.  
 Claim 20; SEQ ID NO 52425; 103pp; English.  
 The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 sequences. (I) is useful as hybridisation probes, polymerase chain  
 reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 and in recombinant production of (II). The polynucleotides are also used  
 in diagnostics as expressed sequence tags for identifying expressed  
 genes. (I) is useful in gene therapy techniques to restore normal  
 activity of (II) or to treat disease states involving (II). (II) is  
 useful for generating antibodies against it, detecting or quantitating a  
 polypeptide in tissue, as molecular weight markers and as a food  
 supplement. (II) and its binding partners are useful in medical imaging  
 of sites expressing (II). (I) and (II) are useful for treating disorders  
 involving aberrant protein expression or biological activity. The  
 polypeptide and polynucleotide sequences have applications in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits to assess biodiversity  
 and to produce other types of data and products dependent on DNA and  
 amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 amino acid sequences of the invention. Note: The sequence data for this  
 patent did not appear in the printed specification, but was obtained in  
 electronic format directly from WIPO at  
 ftp.wipo.int/pub/published\_pct\_sequences  
 SQ Sequence 19 AA;  
 Query Match 1.5%; Score 83; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MAPPTGVLSLLLVTTIA 18  
 Db 1 MAPPTGVLSLLLVTTIA 18  
 RESULT 9  
 ADR00854  
 ID ADR00854 standard; peptide; 24 AA.  
 XX  
 AC ADR00854;  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX

DE Human 254PID6B v.3 peptide SEQ ID NO:265.  
XX 254PID6B; small interfering RNA; siRNA; immune response;  
KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
XX 254PID6B v.3; chromosome 6.  
OS Homo sapiens.  
XX  
XX WO2004067716-A2.  
XX 12-AUG-2004.  
XX  
XX 23-JAN-2004; 2004WO-US001965.  
XX 24-JAN-2003; 2003US-0442526P.  
XX (AGEN-) AGENSYS INC.  
XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
PI Perez-Villar JJ, Faris M;  
XX WPI; 2004-580991/56.  
XX  
XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX  
XX Example 13; SEQ ID NO 265; 345pp; English.  
XX  
XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (p1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of p1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of p1; (5)  
CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of p1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of p1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of p1. 254PID6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents a  
CC human 254PID6B v.3 peptide, which is used in the exemplification of the  
CC present invention. The human 254PID6B gene is located on chromosome 6p22.  
XX  
XX Sequence 24 AA;  
Query Match 1.4%; Score 79; DB 8; Length 24;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 20 CARKQCSGRTYSN 33  
Db 11 CARKQCSGRTYSN 24  
|||||  
|||||  
RESULT 10  
ADR00727

ADR00727 standard; peptide; 9 AA.  
AC ADR00727;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
XX Human 254PID6B peptide SEQ ID NO:138.  
DE  
XX  
XX 254PID6B; small interfering RNA; siRNA; immune response;  
KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
XX chromosome 6.  
XX  
XX Homo sapiens.  
OS  
XX WO2004067716-A2.  
PN  
XX 12-AUG-2004.  
PD  
XX  
XX 23-JAN-2004; 2004WO-US001965.  
PF  
XX 24-JAN-2003; 2003US-0442526P.  
PR  
XX (AGEN-) AGENSYS INC.  
PA  
XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
PI Perez-Villar JJ, Faris M;  
XX WPI; 2004-580991/56.  
DR  
XX  
XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX  
XX Example 9; SEQ ID NO 138; 345pp; English.  
XX  
XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (p1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of p1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of p1; (5)  
CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of p1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of p1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of p1. 254PID6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents a  
CC human 254PID6B peptide, which is used in the exemplification of the  
CC present invention. The human 254PID6B gene is located on chromosome 6p22.  
XX  
XX Sequence 9 AA;  
Query Match 1.1%; Score 63; DB 8; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 946 WDGESNCEW 954





CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254P1D6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 XX  
 SQ Sequence 9 AA;

Query Match 1.1%; Score 59; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 RCICSHLWM 937

Db 1 RCICSHLWM 9

RESULT 15  
 ADNR0838  
 ID ADR00838 standard; peptide; 9 AA.

XX AC ADR00838;

XX DT 04-NOV-2004 (first entry)

XX DE Human 254P1D6B peptide SEQ ID NO:249.

XX KW 254P1D6B; small interfering RNA; siRNA; immune response;  
 KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 XX chromosome 6.

XX OS Homo sapiens.

XX PN WO2004067716-A2.

XX PD 12-AUG-2004.

XX PF 23-JAN-2004; 2004WO-US001965.

XX PR 24-JAN-2003; 2003US-0442526P.

XX PA (AGEN-) AGENSYS INC.

XX PI Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;

XX PI Perez-Villar JJ, Paris M;

XX DR WPI; 2004-580991/56.

XX PT New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX PS Example 9; SEQ ID NO 249; 345pp; English.

XX CC The present invention describes a 254P1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID

CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254P1D6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 XX  
 SQ Sequence 9 AA;

Query Match 1.1%; Score 59; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 654 GIVFYHWEH 662

Db 1 GIVFYHWEH 9

Search completed: October 12, 2005, 10:29:17  
 Job time : 75 secs





254p1d6B-related polynucleotide; (6) a composition that modulates the status of a cell that expresses a protein of Pl; (7) a pharmaceutical composition that comprises the composition described above in a human unit dose form; (8) an antibody or its fragment, which is monoclonal; (9) a non-human transgenic animal that produces the antibody; (10) a hybridoma that produces the antibody; (11) delivering a cytotoxic agent or a diagnostic agent to a cell that expresses the protein of Pl; and (12) inhibiting growth, reproduction or survival of cancer cells that express the protein of Pl. 254p1d6B has cytostatic activity, and can be used in gene therapy. The compositions, molecules and methods are useful for treating and detecting cancer. The present sequence represents the human 254p1d6B v.5, which is used in the exemplification of the present invention. The human 254p1d6B gene is located on chromosome 6p22.

XX Sequence 1072 AA;

Query Match 99.9%; Score 5576; DB 8; Length 1072;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAPPTGVLSSLLLVTTIAGCARKCCSEGRYTSNAVISPNLETTIRMRVSHVTPVVDCTAA 60  
 Db 1 MAPPTGVLSSLLLVTTIAGCARKCCSEGRYTSNAVISPNLETTIRMRVSHVTPVVDCTAA 60  
 Qy 61 CCDSLSSCDLAWFGRGCVLSCPHKENCEPKMGPIRSYLTFLVRPVPORPAQLLDYGDMM 120  
 Db 61 CCDSLSSCDLAWFGRGCVLSCPHKENCEPKMGPIRSYLTFLVRPVPORPAQLLDYGDMM 120  
 Qy 121 LNRGSPGIWGDSPEDIRKDLFLGKDWGLEEMSEYDDYRELEKDLIQLPSGKQEPGRSA 180  
 Db 121 LNRGSPGIWGDSPEDIRKDLFLGKDWGLEEMSEYDDYRELEKDLIQLPSGKQEPGRSA 180  
 Qy 181 EYTWGLLPGSEGAFNSVSGSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240  
 Db 181 EYTWGLLPGSEGAFNSVSGSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240  
 Qy 241 TPSSGEVLEKAKAQLOEQSSNSGKEVIMPSHSLPPASLELSVTVVEKSPVLTVTPGST 300  
 Db 241 TPSSGEVLEKAKAQLOEQSSNSGKEVIMPSHSLPPASLELSVTVVEKSPVLTVTPGST 300  
 Qy 301 EHSIPTPTPSAAPSESTPSSELPISPTTAPRTVKELTVSAGNLIITLPDNEVELKAFVAP 360  
 Db 301 EHSIPTPTPSAAPSESTPSSELPISPTTAPRTVKELTVSAGNLIITLPDNEVELKAFVAP 360  
 Qy 361 APPVETTYNENLISHPTDYQGIKQKHQKTLNLSQLSVGLYVFKVTVSSENAFGEFV 420  
 Db 361 APPVETTYNENLISHPTDYQGIKQKHQKTLNLSQLSVGLYVFKVTVSSENAFGEFV 420  
 Qy 421 NVTVPARRVNLPPVAVVSPQLBELTFLTSALIDGQSSTDDTRIVSYHWEIINGPFTIE 480  
 Db 421 NVTVPARRVNLPPVAVVSPQLBELTFLTSALIDGQSSTDDTRIVSYHWEIINGPFTIE 480  
 Qy 481 KTSVDSPLRLSNLDPGNYSFRLTVTDSGATNSTTAALI VNNAYDVPVANAGNHTIT 540  
 Db 481 KTSVDSPLRLSNLDPGNYSFRLTVTDSGATNSTTAALI VNNAYDVPVANAGNHTIT 540  
 Qy 541 LPQNSITLNGQSSDDHQI VLYEWSLPGSGEGKHVMQGVQTPYLHLSAMQEGDYTFQLK 600  
 Db 541 LPQNSITLNGQSSDDHQI VLYEWSLPGSGEGKHVMQGVQTPYLHLSAMQEGDYTFQLK 600  
 Qy 601 VTDSSRQSTAXVTIVOPENNRPPVAVAGDPDKELI PFVESATLDGSSSSDDHGIVFVHW 660  
 Db 601 VTDSSRQSTAXVTIVOPENNRPPVAVAGDPDKELI PFVESATLDGSSSSDDHGIVFVHW 660  
 Qy 661 EHVGRPSAVENIDKATATVTLGVGTHYHRLTVKQDQGLSSSTLTVAVKKNNSPPR 720  
 Db 661 EHVGRPSAVENIDKATATVTLGVGTHYHRLTVKQDQGLSSSTLTVAVKKNNSPPR 720  
 Qy 721 ARAGGRHVLLPNNSITLDGSRSTDDQIRVSYLWIRDCQSPAAGDVIDGSDHVALQLTN 780  
 Db 721 ARAGGRHVLLPNNSITLDGSRSTDDQIRVSYLWIRDCQSPAAGDVIDGSDHVALQLTN 780  
 Qy 781 LVEGVYTHLVRVDSQGASDTRATVEVQDPDRKSGLVLTQLQVGQLTQRKDTLVRQ 840

Db 781 LVEGVYTHLVRVDSQGASDTRATVEVQDPDRKSGLVLTQLQVGQLTQRKDTLVRQ 840  
 Qy 841 LAVLLNVLDSIDIKVKIRAHSDLSSTVIVFYVQSRPPFKVLAABVARNLHRLSKEKADF 900  
 Db 841 LAVLLNVLDSIDIKVKIRAHSDLSSTVIVFYVQSRPPFKVLAABVARNLHRLSKEKADF 900  
 Qy 901 LIPKVLAVDTAGCLLKCSGHCHCDPLTKRCICSHLWENLIQRYWIDGESNCESIFVVT 960  
 Db 901 LIPKVLAVDTAGCLLKCSGHCHCDPLTKRCICSHLWENLIQRYWIDGESNCESIFVVT 960  
 Qy 961 VLAFTLVLGTGFTWLCICCCCKRQKTKIRKTKYTTILDNNDEQERMELEPKYGIKHRST 1020  
 Db 961 VLAFTLVLGTGFTWLCICCCCKRQKTKIRKTKYTTILDNNDEQERMELEPKYGIKHRST 1020  
 Qy 1021 EHNSLWVSESEFSDQDTTIFSRKMERGNPKVMNGSIRNGASFYSYCKDR 1072  
 Db 1021 EHNSLWVSESEFSDQDTTIFSRKMERGNPKVMNGSIRNGASFYSYCKDR 1072

RESULT 3  
 ADDR00601  
 ID ADDR00601 standard; protein; 1072 AA.  
 XX ADDR00601;  
 AC  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE Human 254p1d6B v.6 protein SEQ ID NO:12.  
 XX  
 KW 254p1d6B; small interfering RNA; siRNA; immune response;  
 KW 254p1d6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254p1d6B v.6; chromosome 6.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004067716-A2.  
 XX  
 PD 12-AUG-2004.  
 XX  
 PF 23-JAN-2004; 2004WO-US001965.  
 XX  
 PR 24-JAN-2003; 2003US-0442526P.  
 XX  
 PA (AGEN-) AGENSYS INC.  
 XX  
 PI Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
 PI Perez-Villar JJ, Paris M;  
 XX  
 DR WPI; 2004-580991/56.  
 XX  
 PT New 254p1d6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254p1d6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.  
 XX  
 PS Example 1; SEQ ID NO 12; 345pp; English.  
 XX  
 CC The present invention describes a 254p1d6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254p1d6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (Pl, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of Pl; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of Pl; (5)

detecting, in a sample, the presence of a 254P1D6B-related protein or a 254P1D6B-related polynucleotide; (6) a composition that modulates the status of a cell that expresses a protein of P1; (7) a pharmaceutical composition that comprises the composition described above in a human unit dose form; (8) an antibody or its fragment, which is monoclonal; (9) a non-human transgenic animal that produces the antibody; (10) a hybridoma that produces the antibody; (11) delivering a cytotoxic agent or a diagnostic agent to a cell that expresses the protein of P1; and (12) inhibiting growth, reproduction or survival of cancer cells that express the protein of P1. 254P1D6B has cytostatic activity, and can be used in gene therapy. The compositions, molecules and methods are useful for treating and detecting cancer. The present sequence represents the human 254P1D6B v.6, which is used in the exemplification of the present invention. The human 254P1D6B gene is located on chromosome 6p22.

Query Match	99.9%;	Score 5574;	DB 8;	Length 1072;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 1069;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
Qy	1	MAPPTGYLSLLLLVTTIAGCARKCOCSEGRYTSNAVISPNETTRIMRVSHFTFPVVDCTAA	60	
Db	1	MAPPTGYLSLLLLVTTIAGCARKCOCSEGRYTSNAVISPNETTRIMRVSHFTFPVVDCTAA	60	
Qy	61	CCDLSSCDLAWWFEGRCYLVSCPHKENCEPKMGPIRSYLTFLVLRPVQRPQAQLLDYDGM	120	
Db	61	CCDLSSCDLAWWFEGRCYLVSCPHKENCEPKMGPIRSYLTFLVLRPVQRPQAQLLDYDGM	120	
Qy	121	LNRGSPGIWGDSPEDIRKDLFLGKWGLEEMSEYDDYRELEKDLLQPSGKQEPGRSA	180	
Db	121	LNRGSPGIWGDSPEDIRKDLFLGKWGLEEMSEYADYRELEKDLLQPSGKQEPGRSA	180	
Qy	181	EYTDWGLLLPGSEGAFNSGVGDSPAVPAETQODPELHYLINESASTAPAKLPERSVLLPLPT	240	
Db	181	EYTDWGLLLPGSEGAFNSGVGDSPAVPAETQODPELHYLINESASTAPAKLPERSVLLPLPT	240	
Qy	241	TPSSGEVLEKEKASQLQEQSSNSGKEVIMPSHSLPPASLBLSVTVVEKSPVLTVTPGST	300	
Db	241	TPSSGEVLEKEKASQLQEQSSNSGKEVIMPSHSLPPASLBLSVTVVEKSPVLTVTPGST	300	
Qy	301	EHSIPTPTTSAAPSESTPSELIPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP	360	
Db	301	EHSIPTPTTSAAPSESTPSELIPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP	360	
Qy	361	APPVETTYNYEWNLLISHPTDYQGEIKQGHKQHTLNLSQLSVGLYVFKVTVSSENAFGEFV	420	
Db	361	APPVETTYNYEWNLLISHPTDYQGEIKQGHKQHTLNLSQLSVGLYVFKVTVSSENAFGEFV	420	
Qy	421	NVTVKPARRVNLPPVAVVSPQQLBELTLPILTSALIDGQSOTDDTETIVSYHWEINGPFTEE	480	
Db	421	NVTVKPARRVNLPPVAVVSPQQLBELTLPILTSALIDGQSOTDDTETIVSYHWEINGPFTEE	480	
Qy	481	KTSVDSVPLRLNSLNDPGNYSFRLTVTDSGDGATNSTTAALIVNNAVDYPPVNAAGPNHIT	540	
Db	481	KTSVDSVPLRLNSLNDPGNYSFRLTVTDSGDGATNSTTAALIVNNAVDYPPVNAAGPNHIT	540	
Qy	541	LPQNSITLNGNQSSDDHQI VLYEWSLGSFGSEGHVVMQGVQTPYLHLSAMQEGDYTFOLK	600	
Db	541	LPQNSITLNGNQSSDDHQI VLYEWSLGSFGSEGHVVMQGVQTPYLHLSAMQEGDYTFOLK	600	
Qy	601	VTDSSRQOSTAXVTVI VQOPENRPPVAVAGDPDKELI PFVESATLDGSSSSDDHGIVFVYH	660	
Db	601	VTDSSRQOSTAXVTVI VQOPENRPPVAVAGDPDKELI PFVESATLDGSSSSDDHGIVFVYH	660	
Qy	661	EHVGPSPAVEMENDKATATVTLQGVGYHYHRLTVKDCQGLSSTSTLTAVKCKENNSPPR	720	
Db	661	EHVGPSPAVEMENDKATATVTLQGVGYHYHRLTVKDCQGLSSTSTLTAVKCKENNSPPR	720	
Qy	721	ARAGRHVLLVLPNNSITLDGSRSTDDORIVSYLWIRTDGQSPAAGDVIDGSDHSVALQUTN	780	
Db	721	ARAGRHVLLVLPNNSITLDGSRSTDDORIVSYLWIRTDGQSPAAGDVIDGSDHSVALQUTN	780	

Qy	781	LVEGVYTHLRVDTSDQASGSDTDTATVEVQDPDRKSGLIVELTLQVGVQLTEQRKDTLVRQ	840
Db	781	LVEGVYTHLRVDTSDQASGSDTDTATVEVQDPDRKSGLIVELTLQVGVQLTEQRKDTLVRQ	840
Qy	841	LAVLLNVLSDSDIKVKIRAHSDLSTVIVFVQSRPPPKVLKAAEVARNLHMRLSKEKADF	900
Db	841	LAVLLNVLSDSDIKVKIRAHSDLSTVIVFVQSRPPPKVLKAAEVARNLHMRLSKEKADF	900
Qy	901	LLFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWENLIQRYIWDGESNCEWSIFYVT	960
Db	901	LLFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWENLIQRYIWDGESNCEWSIFYVT	960
Qy	961	VLAFTLLVLTCGGFTWLICICCCCKQRKTIRKTKYTILDMQDQERMELRPKYGIKHRST	1020
Db	961	VLAFTLLVLTCGGFTWLICICCCCKQRKTIRKTKYTILDMQDQERMELRPKYGIKHRST	1020
Qy	1021	EHNSLMVSESEFSDQDTIFSEKMERGNPKVSMNGSIRNGASFYSYCSKDR	1072
Db	1021	EHNSLMVSESEFSDQDTIFSEKMERGNPKVSMNGSIRNGASFYSYCSKDR	1072
RESULT 4			
AD	ADR00597	ID ADR00597 standard; protein; 1072 AA.	
XX	XX	AC	ADR00597;
XX	XX	AC	
DT	DT	04-NOV-2004 (first entry)	
XX	XX	Human 254P1D6B v.1 clone LCP-3 protein SEQ ID NO:8.	
DE	DE		
XX	XX	254P1D6B; small interfering RNA; siRNA; immune response;	
KW	KW	254P1D6B-related protein; cytosolic; gene therapy; cancer; human;	
KW	KW	254P1D6B v.1 clone LCP-3; chromosome 6.	
XX	XX		
OS	OS	Homo sapiens.	
XX	XX		
PN	PN	WO2004067716-A2.	
XX	XX		
PD	PD	12-AUG-2004.	
XX	XX		
PF	PF	23-JAN-2004; 2004WO-US0001965.	
XX	XX		
PR	PR	24-JAN-2003; 2003US-0442526P.	
XX	XX		
PA	PA	(AGEN-) AGENSYS INC.	
XX	XX		
PI	PI	Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;	
PI	PI	Perez-Villar JJ, Paris M;	
XX	XX		
DR	DR	WPI; 2004-580991/56.	
XX	XX		
PT	PT	New 254P1D6B siRNA composition comprising a double stranded siRNA that	
PT	PT	corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B	
PT	PT	protein or corresponds to a subsequence of the ORF, useful for detecting	
PT	PT	and treating cancer.	
XX	XX		
PS	PS	Example 1; SEQ ID NO 8; 345pp; English.	
XX	XX		
CC	CC	The present invention describes a 254P1D6B small interfering RNA (siRNA)	
CC	CC	composition that comprises a double stranded siRNA that corresponds to	
CC	CC	the nucleic acid open reading frame (ORF) sequence which encodes the	
CC	CC	254P1D6B protein, or corresponds to a subsequence of the ORF, where the	
CC	CC	double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous	
CC	CC	nucleotides in length. Also described: (1) a composition that comprises,	
CC	CC	consists essentially of, or consists of a peptide of eight, nine, ten, or	
CC	CC	eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID	
CC	CC	NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in	
CC	CC	any of the 42 lists of peptides, given in the specification, or a protein	
CC	CC	that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or	
CC	CC	identical to an entire amino acid sequence of P1; (2) a polynucleotide	
CC	CC	that encodes the protein; (3) a composition comprising a polynucleotide	
CC	CC	that is a full complement of the polynucleotide described above; (4)	

CC generating a mammalian immune response directed to the protein of P1; (5)  
CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of P1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents the  
CC human 254P1D6B v.1 clone LCP-3, which is used in the exemplification of  
CC the present invention. The human 254P1D6B gene is located on chromosome  
CC 6p22.  
XX  
XX  
SQ Sequence 1072 AA;  
Query Match 99.9%; Score 5574; DB 8; Length 1072;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 MAPPTGVLSLLLVTTIAGCARKQCSEGRYTSNAVISPNLETTIRMRVSHTFPPVVDCTAA 60  
Db 1 MAPPTGVLSLLLVTTIAGCARKQCSEGRYTSNAVISPNLETTIRMRVSHTFPPVVDCTAA 60  
Qy 61 CCDSLSSCDLAWFEGRCYLVSCPHKENCEPKMGPIRSYLTFLVRPQVPAQLLDYGDMM 120  
Db 61 CCDSLSSCDLAWFEGRCYLVSCPHKENCEPKMGPIRSYLTFLVRPQVPAQLLDYGDMM 120  
Qy 121 LNRGSPGSGWDSPEIRKDLXFGKOWGLEMESEYDDYELKDLQPSGKQEPGRSA 180  
Db 121 LNRGSPGSGWDSPEIRKDLXFGKOWGLEMESEYDDYELKDLQPSGKQEPGRSA 180  
Qy 181 EYTDWGLLPGEAGFNSVGVSDPAPVPAETQODPELHLYNESASTPAPKLPERSVLLPLPT 240  
Db 181 EYTDWGLLPGEAGFNSVGVSDPAPVPAETQODPELHLYNESASTPAPKLPERSVLLPLPT 240  
Qy 241 TPSSGEVLEKEKASQLOQSSNSGKEVLMPSHLSPPASLELSSVTVKSPVLTVTGST 300  
Db 241 TPSSGEVLEKEKASQLOQSSNSGKEVLMPSHLSPPASLELSSVTVKSPVLTVTGST 300  
Qy 301 EHSIPTPTSAPESTPSELPIPTTAPRTVKELTVSAGNLIITLPDNEVELKAFVAP 360  
Db 301 EHSIPTPTSAPESTPSELPIPTTAPRTVKELTVSAGNLIITLPDNEVELKAFVAP 360  
Qy 361 APPVETTYEWNLIISHPTDYQGEIKQGHKOTLNLSQLSVGLYVFKVTSSENAFGGFV 420  
Db 361 APPVETTYEWNLIISHPTDYQGEIKQGHKOTLNLSQLSVGLYVFKVTSSENAFGGFV 420  
Qy 421 NVTVKPARRVNLPPVAVVSPQQLTLPLTSALIDGSGSTDDTEIVSYHWEENGINEPTIEE 480  
Db 421 NVTVKPARRVNLPPVAVVSPQQLTLPLTSALIDGSGSTDDTEIVSYHWEENGINEPTIEE 480  
Qy 481 KTSVDSPLVRLNSLDPGNYSFRLITVDSGATNSTTAALI VNNADVPPVANAGNHTIT 540  
Db 481 KTSVDSPLVRLNSLDPGNYSFRLITVDSGATNSTTAALI VNNADVPPVANAGNHTIT 540  
Qy 541 LPQNSITLNGQSSDDHQIVLYEWSLPGSGKGVVMQGVQTPVYLHLSAMQEGDYTFOLK 600  
Db 541 LPQNSITLNGQSSDDHQIVLYEWSLPGSGKGVVMQGVQTPVYLHLSAMQEGDYTFOLK 600  
Qy 601 VTDSRQOSTAXVTIVOPENNRPPVAVAGDPDKELIPFVEGATLDGSSSSDDHGIVFVHW 660  
Db 601 VTDSRQOSTAXVTIVOPENNRPPVAVAGDPDKELIPFVEGATLDGSSSSDDHGIVFVHW 660  
Qy 661 EHVGPSSAVEMENIDKAIATVTLQVGYTHFRLTVKDOQGLSSSTLTVAVKKNNSPPR 720  
Db 661 EHVGPSSAVEMENIDKAIATVTLQVGYTHFRLTVKDOQGLSSSTLTVAVKKNNSPPR 720  
Qy 721 ARAGGRHVLPNNISITLDGSRSTDQRIYVYLWIRDOQSPAAGDVIDGSHSVALQLTN 780  
Db 721 ARAGGRHVLPNNISITLDGSRSTDQRIYVYLWIRDOQSPAAGDVIDGSHSVALQLTN 780

Db 721 ARAGGRHVLPNNISITLDGSRSTDQRIYVYLWIRDOQSPAAGDVIDGSHSVALQLTN 780  
Qy 781 LVEGVYTFHLRVTDQSASGSDTDTATVEVQDPDKRSLGVLVELTLQVGVGQLTQQRKDTLVRQ 840  
Db 781 LVEGVYTFHLRVTDQSASGSDTDTATVEVQDPDKRSLGVLVELTLQVGVGQLTQQRKDTLVRQ 840  
Qy 841 LAVLLNVLDSDIKVKIRAHSDLSSTVIVFYVQSRPPFKVLKAAEVARNLHMLRSLKEKADF 900  
Db 841 LAVLLNVLDSDIKVKIRAHSDLSSTVIVFYVQSRPPFKVLKAAEVARNLHMLRSLKEKADF 900  
Qy 901 LLFKVLVDTAGCLLKCSGHGHCPLTKRCICSHLWMENLIQRYIWDGESNCEWSIFVYT 960  
Db 901 LLFKVLVDTAGCLLKCSGHGHCPLTKRCICSHLWMENLIQRYIWDGESNCEWSIFVYT 960  
Qy 961 VLAFTLVLVGGFTWLCICCCKQKRTKIRKTKYTTILDNDDEQERMELRPKYGIKHRST 1020  
Db 961 VLAFTLVLVGGFTWLCICCCKQKRTKIRKTKYTTILDNDDEQERMELRPKYGIKHRST 1020  
Qy 1021 EHNSSLMVSESEPDSDQDTTIFSRKMERGNPKVSMNGSIRNGASFYSYCKDR 1072  
Db 1021 EHNSSLMVSESEPDSDQDTTIFSRKMERGNPKVSMNGSIRNGASFYSYCKDR 1072  
RESULT 5  
ADRO0598  
ID ADRO0598 standard; protein; 1072 AA.  
XX ADRO0598;  
XX AC ADRO0598;  
XX DT 04-NOV-2004 (first entry)  
XX DE Human 254P1D6B v.2 protein SEQ ID NO:9.  
XX KW 254P1D6B; small interfering RNA; siRNA; immune response;  
XX KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
XX OS Homo sapiens.  
XX PN WO2004067716-A2.  
XX PD 12-AUG-2004.  
XX PF 23-JAN-2004; 2004WO-US001965.  
XX PR 24-JAN-2003; 2003US-0442526P.  
XX PA (AGEN-) AGENSYS INC.  
XX PI Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
PI Perez-Villar JJ, Paris M;  
XX DR WPI; 2004-580991/56.  
XX PT New 254P1D6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX Example 1; SEQ ID NO 9; 345pp; English.  
XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of P1; (2) a polynucleotide

cc that encodes the protein; (3) a composition comprising a polynucleotide  
cc that is a full complement of the polynucleotide described above; (4)  
cc generating a mammalian immune response directed to the protein of P1; (5)  
cc detecting, in a sample, the presence of a 254p1d6b-related protein or a  
cc 254p1d6b-related polynucleotide; (6) a composition that modulates the  
cc status of a cell that expresses a protein of P1; (7) a pharmaceutical  
cc composition that comprises the composition described above in a human  
cc unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
cc a non-human transgenic animal that produces the antibody; (10) a  
cc hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
cc or a diagnostic agent to a cell that expresses the protein of P1; and  
cc (12) inhibiting growth, reproduction or survival of cancer cells that  
cc express the protein of P1. 254p1d6b has cytostatic activity, and can be  
cc used in gene therapy. The compositions, molecules and methods are useful  
cc for treating and detecting cancer. The present sequence represents the  
cc human 254p1d6b v.2, which is used in the exemplification of the present  
cc invention. The human 254p1d6b gene is located on chromosome 6p22.  
xx  
SQ Sequence 1072 AA;  
Query Match 99.9%; Score 5574; DB 8; Length 1072;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MAPPTGVLSLLLVTTIAGCARQCSEGRYTSNAVISPNLETRIMRVSHTPPVVDCTAA 60  
Db 1 MAPPTGVLSLLLVTTIAGCARQCSEGRYTSNAVISPNLETRIMRVSHTPPVVDCTAA 60  
QY 61 CCLSSCDLAWFEGRCYLVSPHKENCPEKMGPIRSYLTFLVPVQRPQAOLLYDGM 120  
Db 61 CCLSSCDLAWFEGRCYLVSPHKENCPEKMGPIRSYLTFLVPVQRPQAOLLYDGM 120  
QY 121 LNRGSPSGTWGSDPEDIRKDLFLGKWGLEWSEYDDYREXDLQPSKQPRGSA 180  
Db 121 LNRGSPSGTWGSDPEDIRKDLFLGKWGLEWSEYDDYREXDLQPSKQPRGSA 180  
QY 181 EYTDWGLLPGSGAFNSVGDSPAPPAETQQQPELHYLNESASTAPKLPERSVLLPPT 240  
Db 181 EYTDWGLLPGSGAFNSVGDSPAPPAETQQQPELHYLNESASTAPKLPERSVLLPPT 240  
QY 241 TPSSGEVLKESAKSQLOEQSSNSGKEVLMPSHSLPPASLELSVTVVEKSPVLTTPGST 300  
Db 241 TPSSGEVLKESAKSQLOEQSSNSGKEVLMPSHSLPPASLELSVTVVEKSPVLTTPGST 300  
QY 301 EHSIPTPTSAAPSESTPSELPISTTAPRTVKELTVSAGDNLITLPDNEVELKAFVAP 360  
Db 301 EHSIPTPTSAAPSESTPSELPISTTAPRTVKELTVSAGDNLITLPDNEVELKAFVAP 360  
QY 361 APPVETTYNENWNLISHPTDYOGEIKQGHKQTLNLSQLSVGLVYVFKVTSSENAFGEFV 420  
Db 361 APPVETTYNENWNLISHPTDYOGEIKQGHKQTLNLSQLSVGLVYVFKVTSSENAFGEFV 420  
QY 421 NVTVPARRVNLPPVAVVSPQLQELTLPITSALIDGQSQTDDTEIVSYHWEERINGPTEE 480  
Db 421 NVTVPARRVNLPPVAVVSPQLQELTLPITSALIDGQSQTDDTEIVSYHWEERINGPTEE 480  
QY 481 KTSVDSPLRLNLDPCNYSFRLTWTDSGATNSTTAAIIVNNAVDPYPVANAGNPHHTT 540  
Db 481 KTSVDSPLRLNLDPCNYSFRLTWTDSGATNSTTAAIIVNNAVDPYPVANAGNPHHTT 540  
QY 541 LPONSTITLNGNSSDDHQIYLVWESLGPSEGHVVMQGVQTPPYLHLSAQMGEDYTFOLK 600  
Db 541 LPONSTITLNGNSSDDHQIYLVWESLGPSEGHVVMQGVQTPPYLHLSAQMGEDYTFOLK 600  
QY 601 VTDSSRQSQSTAXVTIVQPPENRPPVAVAGPKELIFPVESATLDGSSSSDDHGIVFYHW 660  
Db 601 VTDSSRQSQSTAXVTIVQPPENRPPVAVAGPKELIFPVESATLDGSSSSDDHGIVFYHW 660  
QY 661 EHVGRPSAVEMENIDKAIATATVGLQVGTGTHFRITVKDQOGLSSTLTVAVKKNNSPPR 720  
Db 661 EHVGRPSAVEMENIDKAIATATVGLQVGTGTHFRITVKDQOGLSSTLTVAVKKNNSPPR 720  
QY 721 ARAGGRHVLVLPNNSTLDGSRSTDQRIIVSYLWIRDGQSPAAGDIVDGS DHSV ALQLTN 780

Db 721 ARAGGRHVLVLPNNSTLDGSRSTDQRIIVSYLWIRDGQSPAAGDIVDGS DHSV ALQLTN 780  
QY 781 LVEGVYTFHLRVTDQSGASDTDTATVQVQPPRKSGVLVELTLQVGVGQLTQORKDTLVRO 840  
Db 781 LVEGVYTFHLRVTDQSGASDTDTATVQVQPPRKSGVLVELTLQVGVGQLTQORKDTLVRO 840  
QY 841 LAVLNVLDSDIKVKIRAHSDLSLTVIVFYVQSRPPPFKVLKAAEYARNLHMLRSLSEKADF 900  
Db 841 LAVLNVLDSDIKVKIRAHSDLSLTVIVFYVQSRPPPFKVLKAAEYARNLHMLRSLSEKADF 900  
QY 901 LLFKVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCSEWSIFVYT 960  
Db 901 LLFKVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCSEWSIFVYT 960  
QY 961 VLAFTLIVLTGFTWLCICCCCKROKRTKIRKTKYTILDNMDQERMLRPKYGIKHRST 1020  
Db 961 VLAFTLIVLTGFTWLCICCCCKROKRTKIRKTKYTILDNMDQERMLRPKYGIKHRST 1020  
QY 1021 EHNSLMVSESEFSDSDQDTIFSRKMERGNPKVSNMGSIRNGASFSYCSKOR 1072  
Db 1021 EHNSLMVSESEFSDSDQDTIFSRKMERGNPKVSNMGSIRNGASFSYCSKOR 1072  
RESULT 6  
ADRO0848  
ID ADR00848 standard; protein; 1072 AA.  
XX ADR00848;  
XX 04-NOV-2004 (first entry)  
XX Human 254p1d6b v.1 amino acid sequence SEQ ID NO:259.  
XX 254p1d6b; small interfering RNA; siRNA; immune response;  
KW 254p1d6b-related protein; cytostatic; gene therapy; cancer; human;  
KW 254p1d6b v.1; chromosome 6.  
XX Homo sapiens.  
XX WO2004067716-A2.  
XX 12-AUG-2004.  
XX 23-JAN-2004; 2004WO-US001965.  
XX 24-JAN-2003; 2003US-0442526P.  
XX (AGEN-) AGENSYS INC.  
XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PW, Ge W;  
PI Perez-Villar JJ, Paris M;  
XX WPI; 2004-580991/56.  
XX New 254p1d6b siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254p1d6b  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX Example 13; SEQ ID NO 259; 345pp; English.  
PS The present invention describes a 254p1d6b small interfering RNA (siRNA)  
XX composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254p1d6b protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or

identical to an entire amino acid sequence of p1; (2) a polynucleotide that encodes the protein; (3) a composition comprising a polynucleotide that is a full complement of the polynucleotide described above; (4) generating a mammalian immune response directed to the protein of p1; (5) detecting, in a sample, the presence of a 254p1d6B-related protein or a 254p1d6B-related polynucleotide; (6) a composition that modulates the status of a cell that expresses a protein of p1; (7) a pharmaceutical composition that comprises the composition described above in a human unit dose form; (8) an antibody or its fragment, which is monoclonal; (9) a non-human transgenic animal that produces the antibody; (10) a hybridoma that produces the antibody; (11) delivering a cytotoxic agent or a diagnostic agent to a cell that expresses the protein of p1; and (12) inhibiting growth, reproduction or survival of cancer cells that express the protein of p1. 254p1d6B has cytostatic activity, and can be used in gene therapy. The compositions, molecules and methods are useful for treating and detecting cancer. The present sequence represents a human 254p1d6B v.1 amino acid sequence, which is used in the exemplification of the present invention. The human 254p1d6B gene is located on chromosome 6p22.

XX Seq Sequence 1072 AA;

Query Match 99.9%; Score 5574; DB 8; Length 1072;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAPPTGVLSLLLVTTAGCARKQCSSEGRYTSNAVISPNLETTTRIMRVSHSTFPVVDCTAA 60  
Db 1 MAPPTGVLSLLLVTTAGCARKQCSSEGRYTSNAVISPNLETTTRIMRVSHSTFPVVDCTAA 60  
Qy 61 CCDSLSSCDLAWFGRGCVLSVCPHKECEPKMGPIRSYLTFLVRPVRPAQLLDYDGM 120  
Db 61 CCDSLSSCDLAWFGRGCVLSVCPHKECEPKMGPIRSYLTFLVRPVRPAQLLDYDGM 120  
Qy 121 LNRGSPGIMGDSPEIRKOLXFLGKDWGLBEMSEYDDYRELEKDLQLQPSGKQPRGSA 180  
Db 121 LNRGSPGIMGDSPEIRKOLXFLGKDWGLBEMSEYDDYRELEKDLQLQPSGKQPRGSA 180  
Qy 181 EYTDWGLLPSEGAFNSVGDSPVAPPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240  
Db 181 EYTDWGLLPSEGAFNSVGDSPVAPPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240  
Qy 241 TPSSGEVLEKAKAQLOQSSNSGKEVLMPSHSLPPASLELSVTVTEKSPVLTTPGST 300  
Db 241 TPSSGEVLEKAKAQLOQSSNSGKEVLMPSHSLPPASLELSVTVTEKSPVLTTPGST 300  
Qy 301 EHSIPTPTSAAPSESTSELPISPTTAPRTVKELTVSAGNLIITLPDNEVELKAFVAP 360  
Db 301 EHSIPTPTSAAPSESTSELPISPTTAPRTVKELTVSAGNLIITLPDNEVELKAFVAP 360  
Qy 361 APPVETTYNENLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVPKVTVSSENAFGEFV 420  
Db 361 APPVETTYNENLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVPKVTVSSENAFGEFV 420  
Qy 421 NVTYKPARVNLPPVAVVSPQLELTPLTSALIDGSDSTDDTEIVSYHWEIINGPFTIEE 480  
Db 421 NVTYKPARVNLPPVAVVSPQLELTPLTSALIDGSDSTDDTEIVSYHWEIINGPFTIEE 480  
Qy 481 KTSVDSPLVRLSNLDPGNSYFRLTVTDSGDGATNSTTAALIVNNAVDYPPVANAGNHTIT 540  
Db 481 KTSVDSPLVRLSNLDPGNSYFRLTVTDSGDGATNSTTAALIVNNAVDYPPVANAGNHTIT 540  
Qy 541 LPQNSITLNGQSSDDHQIVLYEWSLPGSGEGKHVVMQGVQTPYLHLSAMQEGDYTFOLK 600  
Db 541 LPQNSITLNGQSSDDHQIVLYEWSLPGSGEGKHVVMQGVQTPYLHLSAMQEGDYTFOLK 600  
Qy 601 VTDSRQOSTAXTVIVQPPENRPPVAVAGDPKELIPPVESATLDGSSSSDDHGIVFYHW 660  
Db 601 VTDSRQOSTAXTVIVQPPENRPPVAVAGDPKELIPPVESATLDGSSSSDDHGIVFYHW 660  
Qy 661 EHVGRPSAVENIDKATATVTVGLQVGYHFRFLTVKQOGLSSNSTLTAVAKKENNSPPR 720  
Db 661 EHVGRPSAVENIDKATATVTVGLQVGYHFRFLTVKQOGLSSNSTLTAVAKKENNSPPR 720

Qy 721 ARAGGRHVLVLPNNISITLDGSRSTDDQRIYSYLWIRDOQSPAAGDVIDGSDHVALQLTN 780  
Db 721 ARAGGRHVLVLPNNISITLDGSRSTDDQRIYSYLWIRDOQSPAAGDVIDGSDHVALQLTN 780  
Qy 781 LVEGVYTFHLRVTDQSGASDTDTATVEVQPPPKSLGVELTLQVGVGQLTEQRKDTLVQ 840  
Db 781 LVEGVYTFHLRVTDQSGASDTDTATVEVQPPPKSLGVELTLQVGVGQLTEQRKDTLVQ 840  
Qy 841 LAVLLNVLDSIKVQKTRAHSDLSLTIVFYVQSRPPFKVLKAAEVARNLHRLSKEKADF 900  
Db 841 LAVLLNVLDSIKVQKTRAHSDLSLTIVFYVQSRPPFKVLKAAEVARNLHRLSKEKADF 900  
Qy 901 LLEFVLAVDATAGCLLKCSGHGCHDPLTKRCICSHLWMLNLIQRYIWDGESNCESIFVYT 960  
Db 901 LLEFVLAVDATAGCLLKCSGHGCHDPLTKRCICSHLWMLNLIQRYIWDGESNCESIFVYT 960  
Qy 961 VLAFTLVLVTGGFTWLICCCCKRQRTKIRKTKYTTLDNNDEOBEMELRPKYGIKHRST 1020  
Db 961 VLAFTLVLVTGGFTWLICCCCKRQRTKIRKTKYTTLDNNDEOBEMELRPKYGIKHRST 1020  
Qy 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFYSYCKDR 1072  
Db 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFYSYCKDR 1072

RESULT 7

ADRO0594

ID ADRO0594 standard; protein; 1072 AA.

XX AC ADRO0594;

XX DT 04-NOV-2004 (first entry)

XX DE Human 254p1d6B v.2 protein SEQ ID NO:5.

XX KW 254p1d6B; small interfering RNA; siRNA; immune response;

XX KW 254p1d6B v.2; chromosome 6.

XX OS Homo sapiens.

XX PN WO2004067716-A2.

XX PD 12-AUG-2004.

XX PF 23-JAN-2004; 2004WO-US001965.

XX PR 24-JAN-2003; 2003US-0442526P.

XX PA (AGEN-) AGENSYS INC.

XX PI Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;

XX PI Perez-Villar JJ, Paris M;

XX XX WPI; 2004-580991/56.

XX DR N-ESDB; ADR00593.

XX XX New 254p1d6B siRNA composition comprising a double stranded siRNA that

PT corresponds to the nucleic acid ORF sequence which encodes the 254p1d6B

PT protein or corresponds to a subsequence of the ORF, useful for detecting

PT and treating cancer.

XX Claim 1; SEQ ID NO 5; 345pp; English.

XX The present invention describes a 254p1d6B small interfering RNA (siRNA)

CC composition that comprises a double stranded siRNA that corresponds to

CC the nucleic acid open reading frame (ORF) sequence which encodes the

CC 254p1d6B protein, or corresponds to a subsequence of the ORF, where the

CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous

CC nucleotides in length. Also described: (1) a composition that comprises,

CC consists essentially of, or consists of a peptide of eight, nine, ten, or

CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID

CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of Pl; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of Pl; (5)  
CC detecting, in a sample, the presence of a 254p1d6B-related protein or a  
CC 254p1d6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of Pl; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of Pl; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of Pl. 254p1d6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents the  
CC human 254p1d6B v.2, which is used in the exemplification of the present  
CC invention. The human 254p1d6B gene is located on chromosome 6p22.  
XX  
SQ Sequence 1072 AA;

Query Match 99.9%; Score 5574; DB 8; Length 1072;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAPPTGVLSLLLVTTIAGCARKQCSGRTYNAVISPNETTRIMRVSHTPPVVDCTAA 60  
DB 1 MAPPTGVLSLLLVTTIAGCARKQCSGRTYNAVISPNETTRIMRVSHTPPVVDCTAA 60  
QY 61 CDSLSSCDLAWFEGRCYLVSPCHKENCPEKMGPIRSYLTFLVPQVORPAOLLDYGDMM 120  
DB 61 CDSLSSCDLAWFEGRCYLVSPCHKENCPEKMGPIRSYLTFLVPQVORPAOLLDYGDMM 120  
QY 121 LNRGSPSGIWDGSPEDIRKDLFLGKDWGLEEMSEYXDDYRELEKDLQPSGKQPRGSA 180  
DB 121 LNRGSPSGIWDGSPEDIRKDLFLGKDWGLEEMSEYADYRELEKDLQPSGKQPRGSA 180  
QY 181 EYTDWGLLPGSGAFNSSVGDSPAVPAETQDQPELHYLNESASTAPKLPERSVLLPPT 240  
DB 181 EYTDWGLLPGSGAFNSSVGDSPAVPAETQDQPELHYLNESASTAPKLPERSVLLPPT 240  
QY 241 TSSGEVLEKASQLOEQSSNSGKEVLMPSHSLPPASLELSVTVKESPVLTTPGST 300  
DB 241 TSSGEVLEKASQLOEQSSNSGKEVLMPSHSLPPASLELSVTVKESPVLTTPGST 300  
QY 301 EHSITPTPTSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360  
DB 301 EHSITPTPTSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360  
QY 361 APPVETTYNENLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAGFEGPV 420  
DB 361 APPVETTYNENLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAGFEGPV 420  
QY 421 NVTVPARRVNLPPVAVVSPQLQELTLPITSALIDGSSQSTDDTEIVSYHWEERINGPFTIE 480  
DB 421 NVTVPARRVNLPPVAVVSPQLQELTLPITSALIDGSSQSTDDTEIVSYHWEERINGPFTIE 480  
QY 481 KTSVDSFVLRLNLDLPNGYSFRLTVTSDGATNSTTAAALIVNNAVDPYPPVANAGPNHTIT 540  
DB 481 KTSVDSFVLRLNLDLPNGYSFRLTVTSDGATNSTTAAALIVNNAVDPYPPVANAGPNHTIT 540  
QY 541 LPQNSITLNGNSSDDHQIIVLYEWSLPGSGEGKHVVMQGVQTPPYLHLSAMQSGDYTFQLK 600  
DB 541 LPQNSITLNGNSSDDHQIIVLYEWSLPGSGEGKHVVMQGVQTPPYLHLSAMQSGDYTFQLK 600  
QY 601 VTDSSRQQTAXVTTVLVQPENNRPPVAVAGPKDELIFPVESATLDGSSSDHGHVFFYHW 660  
DB 601 VTDSSRQQTAXVTTVLVQPENNRPPVAVAGPKDELIFPVESATLDGSSSDHGHVFFYHW 660  
QY 661 EHRGSPSAVEMENIDKAIATVTGLQVGYTHFRLTVKDDQGLSSTSTLTVAVKENNSPPR 720

DB 661 EHRGSPSAVEMENIDKAIATVTGLQVGYTHFRLTVKDDQGLSSTSTLTVAVKENNSPPR 720  
QY 721 ARAGGRHVLPNNISITLDGSRSTDDQRIVSYLATRDGQSPAAGDVIDGSDHSHVALQLTN 780  
DB 721 ARAGGRHVLPNNISITLDGSRSTDDQRIVSYLATRDGQSPAAGDVIDGSDHSHVALQLTN 780  
QY 781 LVEGVYTFHLRVTDTSQGASDSTDATVEVQDPDRKSGLVLTQLQGVGQLTEQKDTLVRQ 840  
DB 781 LVEGVYTFHLRVTDTSQGASDSTDATVEVQDPDRKSGLVLTQLQGVGQLTEQKDTLVRQ 840  
QY 841 LAVLLNVLDSDIKVKQIRAHSDLSITVIVFYVQSPPPFKVLKAAEVARNLHMLSEKADP 900  
DB 841 LAVLLNVLDSDIKVKQIRAHSDLSITVIVFYVQSPPPFKVLKAAEVARNLHMLSEKADP 900  
QY 901 LLEFKVLRVDTAGCLLKCSGHGHCPLTKRCICSHLWMENLIQRYIWDGESNCEWSIFYYT 960  
DB 901 LLEFKVLRVDTAGCLLKCSGHGHCPLTKRCICSHLWMENLIQRYIWDGESNCEWSIFYYT 960  
QY 961 VLAFTLIVLTGGFTWLICICCCCKRQKRTKIRKTKYTTILDNMDEQERMBELRPKYGIKHRST 1020  
DB 961 VLAFTLIVLTGGFTWLICICCCCKRQKRTKIRKTKYTTILDNMDEQERMBELRPKYGIKHRST 1020  
QY 1021 EHNSLWVSESEFSDSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKOR 1072  
DB 1021 EHNSLWVSESEFSDSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKOR 1072  
RESULT 8  
ADRO0592  
ID ADR00592 standard; protein; 1072 AA.  
XX  
AC ADR00592;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Human 254p1d6B v.1 clone LCP-3 protein SEQ ID NO:3.  
KW 254p1d6B; small interfering RNA; siRNA; immune response;  
KW 254p1d6B-related protein; cytostatic; gene therapy; cancer; human;  
KW 254p1d6B v.1 clone LCP-3; chromosome 6.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 142 /label= unknown  
FT /label= "encoded by MCC"  
FT Misc-difference 157 /label= unknown  
FT /label= "encoded by KCA"  
FT Misc-difference 612 /label= unknown  
FT /label= "encoded by GTR"  
XX WO2004067716-A2.  
PN  
PD 12-AUG-2004.  
XX  
PF 23-JAN-2004; 2004WO-US001965.  
XX  
PR 24-JAN-2003; 2003US-0442526P.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
PI Perez-Villar JJ, Faris M;  
XX  
DR WPI: 2004-580991/56.  
DR N-PSDB; ADR00591.  
XX  
PT New 254p1d6B siRNA composition comprising a double stranded siRNA that corresponds to the nucleic acid ORF sequence which encodes the 254p1d6B



PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.

Claim 1; SEQ ID NO 3; 345pp; English.

XX The present invention describes a 254p1d6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254p1d6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO.3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of P1; (5)  
CC detecting, in a sample, the presence of a 254p1d6B-related protein or a  
CC 254p1d6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of P1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of P1. 254p1d6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents the  
CC human 254p1d6B v.1 clone LCP-3, which is used in the exemplification of  
CC the present invention. The human 254p1d6B gene is located on chromosome  
CC 6p22.

XX Sequence 1072 AA;

Query Match 99.9%; Score 5574; DB 8; Length 1072;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1072; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPPTGVLSSLLLVTTAGCARKQCEGRTYSNAVISPNLETRIMRVSHTFPPVVDCTAA 60  
DB 1 MAPPTGVLSSLLLVTTAGCARKQCEGRTYSNAVISPNLETRIMRVSHTFPPVVDCTAA 60  
QY 61 CCDSLSSCDLAWFEGRCYLVSCPHKENCEPKKMGPIRSYTLFVLRPVQRPQALLDYGDM 120  
DB 61 CCDSLSSCDLAWFEGRCYLVSCPHKENCEPKKMGPIRSYTLFVLRPVQRPQALLDYGDM 120  
QY 121 LNRGSPGIWGDSPEDIRKOLXFLGKDWGLEMESEYXDDYRELEKDLQLQPSGKQEPGRSA 180  
DB 121 LNRGSPGIWGDSPEDIRKOLXFLGKDWGLEMESEYXDDYRELEKDLQLQPSGKQEPGRSA 180  
QY 181 EYTWGLLPGSEGAFNSSVGSVPAPVASTQDDPELHYLNESASTPAPKLPERSVLLPLPT 240  
DB 181 EYTWGLLPGSEGAFNSSVGSVPAPVASTQDDPELHYLNESASTPAPKLPERSVLLPLPT 240  
QY 241 TPSSGEVLKESKASQLQSQSSNSGKEVLMPSHSLPPASLELSVTVVEKSPVLTTPGST 300  
DB 241 TPSSGEVLKESKASQLQSQSSNSGKEVLMPSHSLPPASLELSVTVVEKSPVLTTPGST 300  
QY 301 EHSIPTPTSAAPSESTSELPISPTTAPRVKELTVSAGDNLITLDPNEVELKAFVAP 360  
DB 301 EHSIPTPTSAAPSESTSELPISPTTAPRVKELTVSAGDNLITLDPNEVELKAFVAP 360  
QY 361 APPVETTYNENWLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVPKVTVSSENAPGEGFV 420  
DB 361 APPVETTYNENWLISHPTDYQGEIKQGHKQTLNLSQLSVGLYVPKVTVSSENAPGEGFV 420  
QY 421 NVTVKPARRVNLPPVAVVSPQLBELTLPLTSALIDGSQSTDDTEIVSYHWEENINGPFIEE 480  
DB 421 NVTVKPARRVNLPPVAVVSPQLBELTLPLTSALIDGSQSTDDTEIVSYHWEENINGPFIEE 480

QY 481 KTSVDSPLRLSNLDPGNYSPFLTVTDSGATNSTTAAALIVNNAVDPYPPVANAGPNHTIT 540  
DB 481 KTSVDSPLRLSNLDPGNYSPFLTVTDSGATNSTTAAALIVNNAVDPYPPVANAGPNHTIT 540  
QY 541 LPQNSITLNGQSSDDHQIIVLYEWSLPGSGEGKHVVMQGVQTPYLHLISAMQEGDYTFOLK 600  
DB 541 LPQNSITLNGQSSDDHQIIVLYEWSLPGSGEGKHVVMQGVQTPYLHLISAMQEGDYTFOLK 600  
QY 601 VTDSRQOSTAXVTVIIVQENNRPPVAVAGPKELIPVESATLDGSSSSDDHGIVFYHW 660  
DB 601 VTDSRQOSTAXVTVIIVQENNRPPVAVAGPKELIPVESATLDGSSSSDDHGIVFYHW 660  
QY 661 EHVRCPSAVEMENIDKAIATVTGLQVGYTHFRLTVKQOGLSSSTSTLTVAVKENNSPPR 720  
DB 661 EHVRCPSAVEMENIDKAIATVTGLQVGYTHFRLTVKQOGLSSSTSTLTVAVKENNSPPR 720  
QY 721 ARAGRHVLPNNISITLDGSRSDTDORIIVSYLWIRGQSPAAGDVIDGSDHVALQLTN 780  
DB 721 ARAGRHVLPNNISITLDGSRSDTDORIIVSYLWIRGQSPAAGDVIDGSDHVALQLTN 780  
QY 781 LVEGVYTFHLRVTDQSGASDSTDTATVEVQDPDKSLVELTLQVGVGQLTEQRKDTLVRQ 840  
DB 781 LVEGVYTFHLRVTDQSGASDSTDTATVEVQDPDKSLVELTLQVGVGQLTEQRKDTLVRQ 840  
QY 841 LAVLLNVLDSDIKVQKIRAHSDLSSTVIIVYVQSRPPPKVLKAAEVARNLHRLSKEKADF 900  
DB 841 LAVLLNVLDSDIKVQKIRAHSDLSSTVIIVYVQSRPPPKVLKAAEVARNLHRLSKEKADF 900  
QY 901 LLFKVLAVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCEWSIFVYT 960  
DB 901 LLFKVLAVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCEWSIFVYT 960  
QY 961 VLAFTLIVLTGGFTWLCTCCCKRQKTKIRKTKYTTILDNMDEQERMELRPKYGIKHRST 1020  
DB 961 VLAFTLIVLTGGFTWLCTCCCKRQKTKIRKTKYTTILDNMDEQERMELRPKYGIKHRST 1020  
QY 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSNMGSIRNGASFSYCSKDR 1072  
DB 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSNMGSIRNGASFSYCSKDR 1072  
RESULT 9  
ADNR00862  
ID ADR00862 standard; protein; 1053 AA.  
XX AC ADR00862;  
XX DT 04-NOV-2004 (first entry)  
XX Human 254p1d6B v.1 protein sequence SEQ ID NO:273.  
DE 254p1d6B; small interfering RNA; siRNA; immune response;  
KW 254p1d6B-related protein; cytosolic; gene therapy; cancer; human;  
KW 254p1d6B v.1; chromosome 6.  
XX OS Homo sapiens.  
XX WO2004067716-A2.  
XX 12-AUG-2004.  
XX 23-JAN-2004; 2004WO-US001965.  
XX 24-JAN-2003; 2003US-0442526P.  
XX (AGEN-) AGENSYS INC.  
XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
PI Perez-Villar JJ, Paris M;  
XX WPI; 2004-580991/56.  
XX

PT New 254PID6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.

PS Example 5; SEQ ID NO 273; 345pp; English.

XX  
CC The present invention describes a 254PID6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of PI; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of PI; (5)  
CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of PI; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of PI; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of PI. 254PID6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents a  
CC human 254PID6B v.1 protein sequence, which is used in the exemplification  
CC of the present invention. The human 254PID6B gene is located on  
CC chromosome 6p22.

XX Sequence 1053 AA;

SQ Query Match 98.3%; Score 5485; DB 8; Length 1053;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 20 CARQCSGRTYSNAVISNLETRIMRVSHTPVVDCTAACDLSSCDLAWFEGRCYL 79  
DB 1 CARQCSGRTYSNAVISNLETRIMRVSHTPVVDCTAACDLSSCDLAWFEGRCYL 60  
QY 80 VSCPHKENCCEPKMGPIRSYLTFLVLPVQRPQALLDYGDMMLNRGSPSGIWDGSPEDIRK 139  
DB 61 VSCPHKENCCEPKMGPIRSYLTFLVLPVQRPQALLDYGDMMLNRGSPSGIWDGSPEDIRK 120  
QY 140 DLXFLKQKGLSEMSYDDYRELEKDLQPSGKQPRGSAEYTDWGLLPGSEGAFNNSV 199  
DB 121 DLPLFLKQKGLSEMSYDDYRELEKDLQPSGKQPRGSAEYTDWGLLPGSEGAFNNSV 180  
QY 200 GDSPPVAPETQDQPELHYLNESASTAPKLPERSVLLPLPTTSSGCEVLEKSKASOLQEQ 259  
DB 181 GDSPPVAPETQDQPELHYLNESASTAPKLPERSVLLPLPTTSSGCEVLEKSKASOLQEQ 240  
QY 260 SNNSSGKEVLMPSHSLPPASLELSVTVKPSVLTVPSTGSTEHSIPTPTSAAPSESTPS 319  
DB 241 SNNSSGKEVLMPSHSLPPASLELSVTVKPSVLTVPSTGSTEHSIPTPTSAAPSESTPS 300  
QY 320 ELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPAPPVETTYNYEWNLIISHT 379  
DB 301 ELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPAPPVETTYNYEWNLIISHT 360  
QY 380 DYOGELKQGHKQTLNLISQVGLVYFKVTVSSNAFEGGVNVTVPARRVNLPPVAVVS 439  
DB 361 DYOGELKQGHKQTLNLISQVGLVYFKVTVSSNAFEGGVNVTVPARRVNLPPVAVVS 420  
QY 440 PQIQELTLPLTSALIDGQSQTDDTEIVSYHWEENGFPIBEKTSVDSVPLRLSLNLDPGNY 499

DB 421 POLQELTLPLTSALIDGQSQTDDTEIVSYHWEENGFPIBEKTSVDSVPLRLSLNLDPGNY 480  
QY 500 SFRLTVTDSGATNTTAAALIVNNAVDPYPPVANAGPNHTITLPQNSITLNGNSSDDHQI 559  
DB 481 SFRLTVTDSGATNTTAAALIVNNAVDPYPPVANAGPNHTITLPQNSITLNGNSSDDHQI 540  
QY 560 VLYEWSLQPGSEGKHVVMQGVQTPVYLHLSAQMGEGDYTFQLKVTSSRQOQSTAXVTVIYQP 619  
DB 541 VLYEWSLQPGSEGKHVVMQGVQTPVYLHLSAQMGEGDYTFQLKVTSSRQOQSTAXVTVIYQP 600  
QY 620 ENRPPVAVAGPDKELIPFVESATLDGSSSDHGHIVFYHWEHVRGSPSAVENIDKAIKA 679  
DB 601 ENRPPVAVAGPDKELIPFVESATLDGSSSDHGHIVFYHWEHVRGSPSAVENIDKAIKA 660  
QY 680 TVTGLQVCTYHFRLTIVKDOQGLSTSTLTAVAKKENSSPPRARAGGRHVLVLPNNSITLD 739  
DB 661 TVTGLQVCTYHFRLTIVKDOQGLSTSTLTAVAKKENSSPPRARAGGRHVLVLPNNSITLD 720  
QY 740 GSRSTDDQRIYSYLWIRDQSPAAAGDVLDGSDHSAVALQLTNLVEGVYTFHLRVTDSSQAS 799  
DB 721 GSRSTDDQRIYSYLWIRDQSPAAAGDVLDGSDHSAVALQLTNLVEGVYTFHLRVTDSSQAS 780  
QY 800 DTDATATVEVQPPRKSGELVELTLQVGVQQLTEQRKDTLVROLAVLLNVLNLDSDIKVKIRA 859  
DB 781 DTDATATVEVQPPRKSGELVELTLQVGVQQLTEQRKDTLVROLAVLLNVLNLDSDIKVKIRA 840  
QY 860 HSDLSLTVFVYVQSRPPPKVLAARVARNLHMLSKKADFLFKVLAVDTAGCLLKCSG 919  
DB 841 HSDLSLTVFVYVQSRPPPKVLAARVARNLHMLSKKADFLFKVLAVDTAGCLLKCSG 900  
QY 920 HGHCDPLTKRCICSHLWMLNIQRYWIDGSENCSEWSIFVYVLAFTLVLVLTGFTWLCIC 979  
DB 901 HGHCDPLTKRCICSHLWMLNIQRYWIDGSENCSEWSIFVYVLAFTLVLVLTGFTWLCIC 960  
QY 980 CCKQKQRTKIRKTKYKTYTILDNMDQERMELEPKYGIKHSSTEHNSSLMVSESEFSDQDT 1039  
DB 961 CCKQKQRTKIRKTKYKTYTILDNMDQERMELEPKYGIKHSSTEHNSSLMVSESEFSDQDT 1020  
QY 1040 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1072  
DB 1021 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1053  
RESULT 10  
ADR00863  
ID ADR00863 standard; protein; 1053 AA.  
XX AC ADR00863;  
XX DT 04-NOV-2004 (first entry)  
XX DE Human 254PID6B v.3 protein sequence SEQ ID NO:274.  
XX KW 254PID6B; small interfering RNA; siRNA; immune response;  
KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
XX OS Homo sapiens.  
XX PN WO2004067716-A2.  
XX PD 12-AUG-2004.  
XX PF 23-JAN-2004; 2004WO-US001965.  
XX PR 24-JAN-2003; 2003US-0442526P.  
XX PA (AGEN-) AGENSYS INC.  
XX PI Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
PI Perez-Villar JJ, Faris M;  
XX

WPI: 2004-580991/56.

New 254P1D6B siRNA composition comprising a double stranded siRNA that corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B protein or corresponds to a subsequence of the ORF, useful for detecting and treating cancer.

Example 5: SEO ID NO 274: 345pp: English.

The present invention describes a 254pD6B small interfering RNA (siRNA) composition that comprises a double stranded siRNA that corresponds to the nucleic acid open reading frame (ORF) sequence which encodes the 254pD6B protein, or corresponds to a subsequence of the ORF, where the double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous nucleotides in length. Also described: (1) a composition that comprises, consists essentially of, or consists of a peptide of eight, nine, ten, or eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in any of the 42 lists of peptides, given in the specification, or a protein that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or identical to an entire amino acid sequence of P1; (2) a polynucleotide that encodes the protein; (3) a composition comprising a polynucleotide that is a full complement of the polynucleotide described above; (4) generating a mammalian immune response directed to the protein of P1; (5) detecting, in a sample, the presence of a 254pD6B-related protein or a 254pD6B-related polynucleotide; (6) a composition that modulates the status of a cell that expresses a protein of P1; (7) a pharmaceutical composition that comprises the composition described above in a human unit dose form; (8) an antibody or its fragment, which is monoclonal; (9) a non-human transgenic animal that produces the antibody; (10) a hybridoma that produces the antibody; (11) delivering a cytotoxic agent or a diagnostic agent to a cell that expresses the protein of P1; and (12) inhibiting growth, reproduction or survival of cancer cells that express the protein of P1. 254pD6B has cytostatic activity, and can be used in gene therapy. The compositions, molecules and methods are useful for treating and detecting cancer. The present sequence represents a human 254pD6B v.3 protein sequence, which is used in the exemplification of the present invention. The human 254pD6B gene is located on chromosome 6p22.

Sequence 1053 AA;

```
Query Match      98.3%; Score 5485; DB 8; Length 1053;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Qy	20	CARKQCSGRTVSNV	ISPNLETTTRIMRVSH	TPVVDCTAACCDLS	SCDLAWWFEGRCYL	79		
Db	1	CARKQCSGRTVSNV	ISPNLETTTRIMRVSH	TPVVDCTAACCDLS	SCDLAWWFEGRCYL	60		
Qy	80	VSCPHKENCEPKMG	PIRSYLTFLVRPVQ	PAQLLDYGDMM	NARGSPSGTWGDS	PDIRK 139		
Db	61	VSCPHKENCEPKMG	PIRSYLTFLVRPVQ	PAQLLDYGDMM	NARGSPSGTWGDS	PDIRK 120		
Qy	140	DLXFLGKDWGLE	EMSEYXDDYRELE	KOLLQPSGKQEP	RGSAEYTDWGL	LPGSEGAENSSV 199		
Db	121	DLXFLGKDWGLE	EMSEYXDDYRELE	KOLLQPSGKQEP	RGSAEYTDWGL	LPGSEGAENSSV 180		
Qy	200	GUSPAPVAETQO	DELPHYLINESA	STPAKPIERSV	LPLPTTPSSGEV	LEKEKASOLOQ 259		
Db	181	GUSPAPVAETQO	DELPHYLINESA	STPAKPIERSV	LPLPTTPSSGEV	LEKEKASOLOQ 240		
Qy	260	SSNSGKEVLPMS	HSHPASILELSS	VTVEKSPVL	TVTTPGSTHESI	PTTPPTSAAPESTPS 319		
Db	241	SSNSGKEVLPMS	HSHPASILELSS	VTVEKSPVL	TVTTPGSTHESI	PTTPPTSAAPESTPS 300		
Qy	320	ELPIPTTPATPR	VKELTVSAGDN	LIITLPDNEVEL	KAFAVAPPVET	TNYEWNLI	ISHTP 379	
Db	301	ELPIPTTPATPR	VKELTVSAGDN	LIITLPDNEVEL	KAFAVAPPVET	TNYEWNLI	ISHTP 360	
Qy	380	DYQGEIKOGHK	QOTLNL	SQSLVGLV	KVFKVTVSS	ENAFGEFV	NVTVKPARVNL	PPVAVVS 439
Db	361	DYQGEIKOGHK	QOTLNL	SQSLVGLV	KVFKVTVSS	ENAFGEFV	NVTVKPARVNL	PPVAVVS 420

PI Perez-Villar JJ, Fakis M;  
XX WPI; 2004-580991/56.  
XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX Example 5; SEQ ID NO 272; 345pp; English.  
XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of P1; (5)  
CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of P1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of P1. 254PID6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents a  
CC human 254PID6B v.3 protein sequence, which is used in the exemplification  
CC of the present invention. The human 254PID6B gene is located on  
CC chromosome 6p22.  
XX Sequence 1063 AA;  
Query Match 98.3%; Score 5485; DB 8; Length 1063;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 20 CARQCSGRTYSNAVISPNLETTIRIMRVSHTPVVDCTAACCDLSSCDLAWWFGRCYL 79  
DB 11 CARQCSGRTYSNAVISPNLETTIRIMRVSHTPVVDCTAACCDLSSCDLAWWFGRCYL 70  
QY 80 VSCPHKNCBPKMGPIRSYLFVLRLPVQPAQLLDYGDMLNRRGSPGSGWSDPEDIRK 139  
DB 71 VSCPHKNCBPKMGPIRSYLFVLRLPVQPAQLLDYGDMLNRRGSPGSGWSDPEDIRK 130  
QY 140 DLXFLGKDWGLEMSYDDYRELEKDLQPSGKQEPGSAEYTDWGLLPGSGAFNSV 199  
DB 131 DLXFLGKDWGLEMSYDDYRELEKDLQPSGKQEPGSAEYTDWGLLPGSGAFNSV 190  
QY 200 GDSPPAVPAETQDPELHYLNESASTPAPKLPLPERSVLLPLPTTPSSGEVLEKEKASQLQ 259  
DB 191 GDSPPAVPAETQDPELHYLNESASTPAPKLPLPERSVLLPLPTTPSSGEVLEKEKASQLQ 250  
QY 260 SSSSGKEVLMPSHSLPPASLELSVTVVEKSPVLTVTPGSTEHSIPTPTSAAPSESTPS 319  
DB 251 SSSSGKEVLMPSHSLPPASLELSVTVVEKSPVLTVTPGSTEHSIPTPTSAAPSESTPS 310  
QY 320 ELPISPTTAPRTVKELTVSAGDNLIITLPDNEVELKAFVAPAPPVETTYNVEWNLISHT 379  
DB 311 ELPISPTTAPRTVKELTVSAGDNLIITLPDNEVELKAFVAPAPPVETTYNVEWNLISHT 370  
QY 380 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTSSENAFGEFVNVTVKPARRVNLPPVAVVS 439

DB 371 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTSSENAFGEFVNVTVKPARRVNLPPVAVVS 430  
QY 440 POLQELTLPPLTSALIDGSSQSTDDTEIYSYHWEENGPIEBEKTSDSPVLRLSNLDPGNY 499  
DB 431 POLQELTLPPLTSALIDGSSQSTDDTEIYSYHWEENGPIEBEKTSDSPVLRLSNLDPGNY 490  
QY 500 SFRILTVDSDGATNSTTAALIVNNAVDPYPPVANAGPNHTITLPQNSITLNGNQSSDDHQI 559  
DB 491 SFRILTVDSDGATNSTTAALIVNNAVDPYPPVANAGPNHTITLPQNSITLNGNQSSDDHQI 550  
QY 560 VLYEWSLPGSGEGKHVVMQGVQTPYLHLSAQMGEDYTPQLKVTSSRQOSTAXVTVIQOP 619  
DB 551 VLYEWSLPGSGEGKHVVMQGVQTPYLHLSAQMGEDYTPQLKVTSSRQOSTAXVTVIQOP 610  
QY 620 ENNPPVAVAGPDKELIFFVESATLDGSSDDHGIIFYHWEHVRGSAVEMENIDKAIA 679  
DB 611 ENNPPVAVAGPDKELIFFVESATLDGSSDDHGIIFYHWEHVRGSAVEMENIDKAIA 670  
QY 680 TVTGLQVGYTHFRILTVDKQOGLSSTLTAVAKKENNSPPRARAGGRHVLVLPNNSITLD 739  
DB 671 TVTGLQVGYTHFRILTVDKQOGLSSTLTAVAKKENNSPPRARAGGRHVLVLPNNSITLD 730  
QY 740 GSRSTDDQRIIVSWIRDPGQSPAAAGDVI DGDSDHVALQLTNLVEGVYTFHLRVTD SQAS 799  
DB 731 GSRSTDDQRIIVSWIRDPGQSPAAAGDVI DGDSDHVALQLTNLVEGVYTFHLRVTD SQAS 790  
QY 800 DTDATATVEVQDPKRSGLVELTLOVGVQQLTEQRKDTLVRQLAVLLNVLDS DIKVQKIRA 859  
DB 791 DTDATATVEVQDPKRSGLVELTLOVGVQQLTEQRKDTLVRQLAVLLNVLDS DIKVQKIRA 850  
QY 860 HSDLSVTIVVQSRPPKVLKAAEVARNLHMLLSKEKADFLFKVLRVDTAGCLLKCSG 919  
DB 851 HSDLSVTIVVQSRPPKVLKAAEVARNLHMLLSKEKADFLFKVLRVDTAGCLLKCSG 910  
QY 920 HGHCDPLTKRCICSHLWMENLIQRYIMDGESNCWSIFVYTVLAFTLVLITGGFTWLCIC 979  
DB 911 HGHCDPLTKRCICSHLWMENLIQRYIMDGESNCWSIFVYTVLAFTLVLITGGFTWLCIC 970  
QY 980 CCKQKQRTKIRKTKYTIILDNMDQERMEIRPKYGIKHRSTHNSSLMVSESEFSDQDT 1039  
DB 971 CCKQKQRTKIRKTKYTIILDNMDQERMEIRPKYGIKHRSTHNSSLMVSESEFSDQDT 1030  
QY 1040 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKOR 1072  
DB 1031 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKOR 1063  
RESULT 12  
ADRO0596  
ID ADR00596 standard; protein; 1063 AA.  
XX ADR00596;  
AC ADR00596;  
XX 04-NOV-2004 (first entry)  
DT 04-NOV-2004 (first entry)  
XX Human 254PID6B v.3 protein SEQ ID NO:7.  
DE 254PID6B; small interfering RNA; siRNA; immune response;  
KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
KW 254PID6B v.3; chromosome 6.  
XX Homo sapiens.  
XX WO2004067716-A2.  
PN 12-AUG-2004.  
XX 23-JAN-2004; 2004WO-US001965.  
PF 24-JAN-2003; 2003US-0442526P.  
PR (AGEN-) AGENSYS INC.  
PA

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge M;  
 PI Perez-Villar JJ, Paris M;  
 XX WPI; 2004-580991/56.  
 DR N-PSDB; ADR00595.  
 XX New 254p1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254p1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.  
 XX Claim 1; SEQ ID NO 7; 345pp; English.  
 PS  
 XX The present invention describes a 254p1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254p1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254p1D6B-related protein or a  
 CC 254p1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254p1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254p1D6B v.3, which is used in the exemplification of the present  
 CC invention. The human 254p1D6B gene is located on chromosome 6p22.  
 XX  
 SQ Sequence 1063 AA;  
 Query Match 98.3%; Score 5485; DB 8; Length 1063;  
 Best Local Similarity 99.7%; Pred No. 0;  
 Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 20 CARQCSGRTYSNAVISPNLETRIMRVSHFTFPVVDCTAACCDLSSCDLAWFEGRCYL 79  
 DB 11 CARQCSGRTYSNAVISPNLETRIMRVSHFTFPVVDCTAACCDLSSCDLAWFEGRCYL 70  
 QY 80 VSCPHKENCEPKMGPIRSYLTFLVLRPVRPAQLLDYGDMLNRGSPSGIWDGSPEDIRK 139  
 DB 71 VSCPHKENCEPKMGPIRSYLTFLVLRPVRPAQLLDYGDMLNRGSPSGIWDGSPEDIRK 130  
 QY 140 DLXFLGKDWGLEEMSEYXDDYRELEKDLQPSGKQEPGSAEYTDWGLLPGEAGFNSSV 199  
 DB 131 DLPLFLGKDWGLEEMSEYSDRELEKDLQPSGKQEPGSAEYTDWGLLPGEAGFNSSV 190  
 QY 200 GDSAPVPAETQODPELHYNESASTPAPKLPERSVLLPLPTTPSSGGEVLEKEKASQLQEQ 259  
 DB 191 GDSAPVPAETQODPELHYNESASTPAPKLPERSVLLPLPTTPSSGGEVLEKEKASQLQEQ 250  
 QY 260 SSNSGKGVLPMSHSLPPASLELSVTVKEKSPVLTPPGSTEHSIPTPTPTSAAPSESTPS 319  
 DB 251 SSNSGKGVLPMSHSLPPASLELSVTVKEKSPVLTPPGSTEHSIPTPTPTSAAPSESTPS 310  
 QY 320 ELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPPVETTYNENWLLISHT 379  
 DB 311 ELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPPVETTYNENWLLISHT 370

QY 380 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSSENAFGEQFVNTVKPARRVNLPRVAVS 439  
 DB 371 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSSENAFGEQFVNTVKPARRVNLPRVAVS 430  
 QY 440 POLQELTLPLTSALIDGQSQTDDTTEIVSYHWEENGPFIEEKTSDVSPVLRLSLNLDRCNY 499  
 DB 431 POLQELTLPLTSALIDGQSQTDDTTEIVSYHWEENGPFIEEKTSDVSPVLRLSLNLDRCNY 490  
 QY 500 SFRLTVTDSGATNSTTAALIVNNAVDPVPPVANAGPNTHTITLQNSITLNGNQSSDDHQI 559  
 DB 491 SFRLTVTDSGATNSTTAALIVNNAVDPVPPVANAGPNTHTITLQNSITLNGNQSSDDHQI 550  
 QY 560 VLYEWSLGPSEGGKHVMQGVQTPYLHLSAMQEGDYTFOLKVTDSRRQOSTAXVTVIQOP 619  
 DB 551 VLYEWSLGPSEGGKHVMQGVQTPYLHLSAMQEGDYTFOLKVTDSRRQOSTAXVTVIQOP 610  
 QY 620 ENNRPPVAVAGDPKELIFPVESATLDGSSSDDDHGIVFYHWEHVRGSPSAVEMENIDKAIA 679  
 DB 611 ENNRPPVAVAGDPKELIFPVESATLDGSSSDDDHGIVFYHWEHVRGSPSAVEMENIDKAIA 670  
 QY 680 TVTGLQVGTYHFRLTVDKQQLSSTLTVAVKKENNSPPRARAGRHVVLVLPNNSITLD 739  
 DB 671 TVTGLQVGTYHFRLTVDKQQLSSTLTVAVKKENNSPPRARAGRHVVLVLPNNSITLD 730  
 QY 740 GSRSTDQRIIVSYLWIRDGQSPAAGDVI DGDHSDHVALQLTNLVSGVYTFHLRVTDGAS 799  
 DB 731 GSRSTDQRIIVSYLWIRDGQSPAAGDVI DGDHSDHVALQLTNLVSGVYTFHLRVTDGAS 790  
 QY 800 DTDATVVEVQDPDKSGELVELTLQVGVQLTQEQKDTLVRQLAVLLNVLSDIKVKQKIRA 859  
 DB 791 DTDATVVEVQDPDKSGELVELTLQVGVQLTQEQKDTLVRQLAVLLNVLSDIKVKQKIRA 850  
 QY 860 HSDLTSTVIVFYVQSRPFKVLKAAEVARNLHMLSKKADFLFKVLAVDTAGCLLKCSG 919  
 DB 851 HSDLTSTVIVFYVQSRPFKVLKAAEVARNLHMLSKKADFLFKVLAVDTAGCLLKCSG 910  
 QY 920 HGHCDDPLTKRCICSHLWENLIQRYIWDGSCNCEWSIFYYVTVLAFTLIVLTGGFTWLCIC 979  
 DB 911 HGHCDDPLTKRCICSHLWENLIQRYIWDGSCNCEWSIFYYVTVLAFTLIVLTGGFTWLCIC 970  
 QY 980 CCKRQKTKIRKTKYTI LNDNMQERMEQLRPKYGIKHRSTEHNSSLMVSESEFSDQDT 1039  
 DB 971 CCKRQKTKIRKTKYTI LNDNMQERMEQLRPKYGIKHRSTEHNSSLMVSESEFSDQDT 1030  
 QY 1040 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1072  
 DB 1031 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1063  
 RESULT 13  
 ADR00599  
 ID ADR00599 standard; protein; 1063 AA.  
 XX ADR00599;  
 AC ADR00599;  
 XX 04-NOV-2004 (first entry)  
 DT  
 XX Human 254p1D6B v.3 protein SEQ ID NO:10.  
 DE  
 XX 254p1D6B; small interfering RNA; siRNA; immune response;  
 KW 254p1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254p1D6B v.3; chromosome 6.  
 XX Homo sapiens.  
 OS  
 XX WO2004067716-A2.  
 PN  
 XX 12-AUG-2004.  
 PD  
 XX 23-JAN-2004; 2004WO-US001965.  
 PF  
 XX 24-JAN-2003; 2003US-0442526P.  
 PR



XX (AGEN-) AGENSYS INC.  
XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eld PM, Ge W;  
PI Perez-Villar JJ, Faris M;  
XX WPI; 2004-580991/56.  
XX New 254p1D6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254p1D6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX Example 1; SEQ ID NO 10; 345bp; English.  
XX The present invention describes a 254p1D6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254p1D6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (Pl, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 homologous or  
CC identical to an entire amino acid sequence of Pl; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of Pl; (5)  
CC detecting, in a sample, the presence of a 254p1D6B-related protein or a  
CC 254p1D6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of Pl; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of Pl; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of Pl. 254p1D6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents the  
CC human 254p1D6B v.3, which is used in the exemplification of the present  
CC invention. The human 254p1D6B gene is located on chromosome 6p22.  
XX Sequence 1063 AA;  
SQ Query Match 98.3%; Score 5485; DB 8; Length 1063;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 20 CARQCSGRTYSNAVISNLETTTRIMRVSHTPVVDCTAACCDLSSCDLAWFGRCYL 79  
Db 11 CARQCSGRTYSNAVISNLETTTRIMRVSHTPVVDCTAACCDLSSCDLAWFGRCYL 70  
QY 80 VSCPHKENCEPKMGPIRSYLVFLPVRQAPQLLDYGDMLNRGSPSGIWDGSDPEDIRK 139  
Db 71 VSCPHKENCEPKMGPIRSYLVFLPVRQAPQLLDYGDMLNRGSPSGIWDGSDPEDIRK 130  
QY 140 DLXFLGKDWGLEMSYDDYRELEKDLLQPSGKQPRGSAEYTDWGLLPGSEGAFNSSV 199  
Db 131 DLXFLGKDWGLEMSYDDYRELEKDLLQPSGKQPRGSAEYTDWGLLPGSEGAFNSSV 190  
QY 200 GDSPAVPAETQDPELHYLNESASTPAPKLERSVLLPLPTPSSEVLEKEKASQLOBQ 259  
Db 191 GDSPAVPAETQDPELHYLNESASTPAPKLERSVLLPLPTPSSEVLEKEKASQLOBQ 250  
QY 260 SNNSSKEVLMPSHSLPPASLELSSVTVEKSPVLTVTPGSTHESIPTPTSAAPSESTPS 319  
Db 251 SNNSSKEVLMPSHSLPPASLELSSVTVEKSPVLTVTPGSTHESIPTPTSAAPSESTPS 310  
QY 320 ELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPPVETTYNYEWNLISHT 379

Db 311 ELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPPVETTYNYEWNLISHT 370  
QY 380 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTYSSENAFGEGFVNVTVKDARRVNLPPVAVS 439  
Db 371 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTYSSENAFGEGFVNVTVKDARRVNLPPVAVS 430  
QY 440 POLQELTLPLTSALIDGSDTDDTEIYSYHWEENGPFIEKTSVDSPLVLSNLDPNY 499  
Db 431 POLQELTLPLTSALIDGSDTDDTEIYSYHWEENGPFIEKTSVDSPLVLSNLDPNY 490  
QY 500 SFRLTVTDSGATNSTTAALIVNNAVDPYPPVANAGPNHTITLPONSTTLNGNQSSDDHOI 559  
Db 491 SFRLTVTDSGATNSTTAALIVNNAVDPYPPVANAGPNHTITLPONSTTLNGNQSSDDHOI 550  
QY 560 VLYEWSLGPSEGHVVMQGVQTPYLHLSAMQEGDYTFOLKVTVDSSRQOSTAXVTIVQP 619  
Db 551 VLYEWSLGPSEGHVVMQGVQTPYLHLSAMQEGDYTFOLKVTVDSSRQOSTAXVTIVQP 610  
QY 620 ENNRPPVAVAGPDKELIFFVESATLDGSSSDDDHGI VFYHWEHVRGSAVEMENIDKAIA 679  
Db 611 ENNRPPVAVAGPDKELIFFVESATLDGSSSDDDHGI VFYHWEHVRGSAVEMENIDKAIA 670  
QY 680 TVTGLQVGTYHFRLTVDKQOGLSTSTLTVAVKKENNSPPRARAGGRHVLVLPNNSITLD 739  
Db 671 TVTGLQVGTYHFRLTVDKQOGLSTSTLTVAVKKENNSPPRARAGGRHVLVLPNNSITLD 730  
QY 740 GSRSTDQRI VSYLWIRDGSPAAGDVI DGDSDHVALQLTNLVEGVYTFHLRVTDSDQAS 799  
Db 731 GSRSTDQRI VSYLWIRDGSPAAGDVI DGDSDHVALQLTNLVEGVYTFHLRVTDSDQAS 790  
QY 800 DTDATATVEVQDPDKRSGLVELTLQVGVQQLTEQRKDTLVRQLAVLLNVLNLDSDIKVKIRA 859  
Db 791 DTDATATVEVQDPDKRSGLVELTLQVGVQQLTEQRKDTLVRQLAVLLNVLNLDSDIKVKIRA 850  
QY 860 HSDLSLTVIVFYVQGRPPFKVLKAAEVARNLHMLRSKEKADFLFKVLAVDVTAGCLLKCSG 919  
Db 851 HSDLSLTVIVFYVQGRPPFKVLKAAEVARNLHMLRSKEKADFLFKVLAVDVTAGCLLKCSG 910  
QY 920 HGHCPLTKRCICSHLWMLNLIQRYIWDGSENCWSIFVYTVLAFTLVLVLTGGFTWLCIC 979  
Db 911 HGHCPLTKRCICSHLWMLNLIQRYIWDGSENCWSIFVYTVLAFTLVLVLTGGFTWLCIC 970  
QY 980 CCKQKRTKIRKTKYITLNDMDQERMELEPKYGIKHSRTEHNSLVMVSESEFSDQDT 1039  
Db 971 CCKQKRTKIRKTKYITLNDMDQERMELEPKYGIKHSRTEHNSLVMVSESEFSDQDT 1030  
QY 1040 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1072  
Db 1031 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1063  
RESULT 14  
ADGI4994  
ID ADGI4994 standard; protein; 978 AA.  
XX  
AC ADGI4994;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Human SECP-27 protein.  
XX  
KW SECP; secreted protein; neuroprotective; relaxant; antithyroid;  
KW antidiabetic; cytosolic; dermatological; immunosuppressive;  
KW antiinflammatory; thyromimetic; antiallergic; cerebroprotective;  
KW gastrointestinal; hepatotropic; nephrotropic; fungicide; protozoacide;  
KW antiparkinsonian; antibacterial; antiparasitic; cardiatic; cardiovascular; anti-HIV;  
KW virucide; uropathic; antirheumatic; cardiant; Parkinson's disease; Alzheimer's; muscular;  
KW neurotic; neurodegenerative; Parkinson's disease; Alzheimer's; muscular;  
KW myotonic dystrophy; catatonia; endocrine; diabetes; leukaemia;  
KW cervical; breast cancer; immunological; scleroderma;  
KW systemic lupus erythematosus; allergy; gastrointestinal; Crohn's; renal;  
KW Goodpasture's syndrome; viral infection; bacterial; fungal; parasitic;  
KW protozoal; helminthic; cardiovascular; atherosclerosis; hepatic;





XX 09-JUL-2003; 2003WO-JP008690.  
 XX 10-JUL-2002; 2002JP-00201856.  
 XX (TAKE ) TAKEDA CHEM IND LTD.  
 XX Matsuzawa Y, Funahashi T, Shimomura C, Furuyama N;  
 PI WPI: 2004-122943/12.  
 XX N-PSDB; ADM36226.  
 XX Mouse membrane and secretory proteins of adipocyte origin and  
 PT polynucleotides encoding them for screening compounds as remedies for  
 PT obesity, diabetes, arteriosclerosis, hypertension and hyperlipemia.  
 XX Claim 41; SEQ ID NO 22; 195pp; Japanese.  
 XX This invention relates to novel membrane proteins that are associated  
 CC with differentiation and/ or metabolic function of adipocytes, in  
 CC particular of mouse origin. Specifically, it refers to the isolated  
 CC nucleic acid molecules encoding all or part of these proteins,  
 CC appropriate antibodies and screening assays useful for the development of  
 CC drug compositions derived thereof. The present invention describes these  
 CC compositions as useful for the treatment of diseases associated with  
 CC abnormalities of adipocyte function, such that they can be used to  
 CC prevent, treat or diagnose obesity, hypertension, hyperlipaemia, diabetes  
 CC and arteriosclerosis. Accordingly, they exhibit anorectic, antidiabetic,  
 CC antiarteriosclerotic, antilipemic and hypotensive activities. This  
 CC polypeptide is a murine membrane protein sequence related to adipocyte  
 CC function of the invention.  
 XX Sequence 1048 AA;  
 SQ  
 Query Match 43.5%; Score 2426.5; DB 8; Length 1048;  
 Best Local Similarity 48.1%; Pred. No. 4.4e-168;  
 Matches 519; Conservative 136; Mismatches 314; Indels 109; Gaps 16;  
 QY 10 SLLLVITAGCARKQCEGRYTSNAVISPNLETRIMRYSHTFPVVDCYAAACCDLSSCDL 69  
 DB 41 SVLWJSTDADEGR--CQCKLYGAGLRTGE-NHURLLAGSLPFFHACRAACCRDSACHA 97  
 QY 70 AMWFEGRYLVSOPHCENCEPKKMGPIRSYLPVLRPVQRAQLLDYDGMMLNRGSPSGI 129  
 DB 98 LNWLEGCWCFQADCKPQSCQPPFTDSSNML-----IIFQKQTTDD 139  
 QY 130 WGDSPEDIRKDLXFLKDWGLER-----MSEYXDDYRELEKOLLQPSGKQEP 177  
 DB 140 LGLLPEDEDFHLLRLG--WGRTSWRRQSLLGAPLTLSPFSSHQSLLRD----- 186  
 QY 178 GSAEYTDWGLLPGEAGFNSSVGDSPAPVAPETQDDPELH-----YLNESASTPAPKLP 230  
 DB 187 --RQKDLVSVPFHGAMQHSKVNHSSEAGALSPTSAEVRKTTIVAGSFYSNHTTQTPWP 244  
 QY 231 ERSVLLPLPTTPSSGGEVLEKEKASQIQEQQSSNSGKEVLMPSHSLPAPLSLSSVTVEKS 290  
 DB 245 KNVSIHPEPS-----EHSSPVSG----- 262  
 QY 291 PVLTVTPGSTEHSIPTTPSAAPSES--TPS-ELPISPTAPR-TVKELTVSAGDNLIT 346  
 DB 263 ---TPQVKSTHSPTDAPLPVAPSYATPTQOASQSTASPHPVVKELWSAGKSVQIT 319  
 QY 347 LPDNEVELKAFAPAPVETTYNVEWNLISHPTDYGKEIKQGHQTLNLSQLSVGLYVFK 406  
 DB 320 LPKNEVQLNAFVLPAPAPGETTYTDWQLITHPTDYSGEVERKHSQSLQSLKPLPGLYEFK 379  
 QY 407 VTVSSNAFGEGFVNVTVKPARVNLPPVAVVSFQELTLPITSALIDGSQSTDDTEIV 466  
 DB 380 VTVDGQNAHGEVNVTVKPEPKRNPVAVVSPQEQEISLPTTSTIIDGSQSTDDDKIV 439  
 QY 467 SYHWEELNGPPIEKEKTSVDSVPLRLNLDGNYSPRLTVDSGATNSTAALIYNNAVD 526  
 DB 440 QYHWEELKGPLREEKISDPTAILKLSKLVPGNVTFSLTVDVSDGATNSTASLTVNKAYD 499

QY 527 YPPVANAGPNHTITLPQNSITLNGNQSSDDHQIIVLYEWSLGGSEGHVNVQGVQTPYLH 586  
 DB 500 YPPVANAGPNQVITLPQNSITLFCNQSTDDHGITSYEWLSLSPSKGKVVEVQGVRTPALQ 559  
 QY 587 LSAMQEGDYTFQLKVTDSRQOSTAKVTIVIQPENRPPVAVAGPDKELIIPVESATLDG 646  
 DB 560 LSAMQEGDYTYQLTVDTTAGQQAQVTVIVQPENRPPQADAGPDKELTLPLVDSTLTDG 619  
 QY 647 SSSDDHGIVFYHWEHVRGSAVEMENIDKAIATVTGLQVGTYHFRLTIVKQOQGLSSTST 706  
 DB 620 SKSTDQRQVSVSYLWEQSRGPDGVQLENANSSVATVTGLQVGTYYFTLTVKQERNLQOSS 679  
 QY 707 LTVAVKKEENSPRARAGRHVLVLPNNISITLDGSRSTDQRIIVSYLWIRDGQSPAAGDV 766  
 DB 680 VNVIVKEINKPPVAKIAGNVVVTLPSTAELDGSRSSDDKGIIVSYLWTRDTSAPAGEV 739  
 QY 767 IDGSDHVALQLTNLVEGVYTFHURVTDQSQASDTDTATVEVQDPDKSLVELTLQVGV 826  
 DB 740 LNHSDHHPVFLSNLVEGTYTFHLKVTDAKESDTRTVEVKDPDKRSLNLEIILDVNV 799  
 QY 827 GOLTEQRKDTLAVROLAVLLNVLDSDIIVQKIRAHSDLSLTVIVYVQSRPPPKVLKAAEVA 886  
 DB 800 SOLTERLKGMLIRQILQVLLGLVDSIIIVQKIQPTTEQSTKMLFFVQNDPPHQLFKGHEVA 859  
 QY 887 RNLHWRLSKEKADFLFKVLAVDVTAGCLLKSGHGHCDPLTKRCITCSHLMENLIQRYIW 946  
 DB 860 AMLKSELQKQKADFLIFALFISTVTCQLNCSDHGHCDSTFKRCVCDPFWMENFIKVQLR 919  
 QY 947 DGENCSEWSIPYTVVLAPTLLVLTGGFTWLCICCCQKQKTKIRKTKYKTYTILDNDEQER 1006  
 DB 920 DGDNSCEWSVLYVIIASFVIVVALGILSWTTTICCCQKQK-GKPKRKSRYKILDATD-QES 977  
 QY 1007 MELRP--KYGIKHSRSTHNSLMVSESEFSDQDTIFSRKMERGNPKVSNMGSIRNG 1062  
 DB 978 LELKPTSRAGSKQKQPTLSSSLMHSELSDDSD-DAIFTWPDREKGLLYGQNGSVFNG 1034

Search completed: October 12, 2005, 10:02:48  
 Job time : 131 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:00:24 ; Search time 128 Seconds  
(without alignments)  
4288.663 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 5580  
Sequence: 1 MAPPTGVLSLLLVTTIAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 03.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result #	Score	Query Match %	Length	DB ID	Description
1	5574	99.9	1109	2 Q9Y4G7	Q9Y4G7 homo sapien
2	4255.5	76.3	1083	2 Q80U39	Q80U39 mus musculu
3	3671	65.8	707	2 Q7L8R6	Q7L8R6 homo sapien
4	2426.5	43.5	1048	2 Q8K135	Q8K135 mus musculu
5	2403.5	43.1	1085	2 Q8BHU7	Q8BHU7 mus musculu
6	2403.5	43.1	1085	2 Q8BHJ2	Q8BHJ2 mus musculu
7	2402	43.0	1049	2 Q8PJU7	Q8PJU7 homo sapien
8	2401	43.0	1049	2 Q8IZAO	Q8IZAO homo sapien
9	2316.5	41.5	946	2 Q8WY25	Q8WY25 homo sapien
10	2256.5	40.4	746	2 Q8NDA0	Q8NDA0 homo sapien
11	2086.5	37.4	691	2 Q9H7V0	Q9H7V0 homo sapien
12	2004	35.9	639	2 Q86J30	Q86J30 homo sapien
13	1675.5	30.0	603	2 Q8VBZ9	Q8VBZ9 mus musculu
14	1454	26.1	684	2 Q8BHR5	Q8BHR5 mus musculu
15	1451.5	26.0	1069	2 Q9VSC9	Q9VSC9 drosophila
16	1440.5	25.8	685	2 Q9H9L2	Q9H9L2 homo sapien
17	1390	24.9	985	2 Q7O8G8	Q7O8G8 anopheles g
18	1358	24.3	491	2 Q8N2B3	Q8N2B3 homo sapien
19	1296.5	23.2	640	2 Q8WY39	Q8WY39 homo sapien
20	931.5	16.7	353	2 Q96IC3	Q96IC3 homo sapien
21	887.5	15.9	695	2 Q7YT23	Q7YT23 drosophila
22	582	10.4	309	2 Q95010	Q95010 homo sapien
23	452.5	8.1	189	2 Q8BUW6	Q8BUW6 homo sapien
24	246.5	4.4	1899	2 Q87RU0	Q87RU0 symbiobacte
25	246	4.4	3441	2 Q89PB9	Q89PB9 bradyrhizob
26	238.5	4.3	5020	2 Q8E9W3	Q8E9W3 shewanella
27	225	4.0	2009	2 Q7UY44	Q7UY44 rhodospirell
28	225	4.0	26926	2 Q8WZB3	Q8WZB3 homo sapien
29	225	4.0	34350	2 Q8WZ42	Q8WZ42 homo sapien
30	224	4.0	26926	2 Q10466	Q10466 homo sapien
31	223.5	4.0	2566	2 Q8TSE7	Q8TSE7 methanosarc

## ALIGNMENTS

### RESULT 1

Q9Y4G7	Q9Y4G7	PRELIMINARY;	PRT;	1109 AA.
AC	Q9Y4G7; Q9UJC8;			
DT	01-NOV-1999 (TrEMBLrel. 12, Created)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	KIAA0319 protein (Fragment).			
GN	Name=KIAA0319;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=97349984; PubMed=9205841;			
RA	Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";			
RL	DNA Res. 4:141-150(1997).			
DR	EMBL; AB002317; BAA20777.2; --			
DR	Genew; HGNC:21580; KIAA0319.			
DR	InterPro; IPR003961; FN III.			
DR	InterPro; IPR000601; PKD.			
DR	SMART; SM00060; FN3; 4.			
DR	SMART; SM00059; PKD; 5.			
DR	PROSITE; PS50093; PKD; 1.			
FT	NON TER 1			
SQ	SEQUENCE 1109 AA; 121961 MW; E39ABFCE6E17BB1C CRC64;			
Query Match 99.9%; Score 5574; DB 2; Length 1109;				
Best Local Similarity 99.7%; Pred. No. 5.8e-265; Mismatches 0; Indels 0; Gaps 0;				
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
Qy	1 MAPPTGVLSLLLVTTIAGCARKKCSGRGYSNAVISPNLETTTRIMRVSHFTFPVVDCTAA 60			
Db	38 MAPPTGVLSLLLVTTIAGCARKKCSGRGYSNAVISPNLETTTRIMRVSHFTFPVVDCTAA 97			
Qy	61 CCILSSCDLAWFGRGCVLVSCHPKNCEPKMGPIRSYLTFLVRPQRPQALLDYGDM 120			
Db	98 CCILSSCDLAWFGRGCVLVSCHPKNCEPKMGPIRSYLTFLVRPQRPQALLDYGDM 157			
Qy	121 LNRGSPSGIGDSDPEDIRKDLXFLGDKDGLFEMSEYKDDYRELEKDLLQPSGKQBPGRSA 180			
Db	158 LNRGSPSGIGDSDPEDIRKDLXFLGDKDGLFEMSEYKDDYRELEKDLLQPSGKQBPGRSA 217			
Qy	181 EYTDWGLLPGSEGAFFNSVGDSPAVPAETQDDPELHYLINESASTPAPKLPERSVLLPLPT 240			
Db	218 EYTDWGLLPGSEGAFFNSVGDSPAVPAETQDDPELHYLINESASTPAPKLPERSVLLPLPT 277			
Qy	241 TPSSGEVLEKEASQLQEQSSNSGKVEVLMPSHSLPPASLESLSSVTVEKSPVLTVPGST 300			

```

Db 278 TPSSGEVLEKASQLOEQSSNSSGKVLMPHSHLPASLELSSVTVKSPVLVTPGST 337
Qy 301 EHSITPPTSAAPSESTSELPISPTTAPRTVKELTVSAGDNLIIITLDPNEVELKAFVAP 360
Db 338 EHSITPPTSAAPSESTSELPISPTTAPRTVKELTVSAGDNLIIITLDPNEVELKAFVAP 397
Qy 361 APPVETTYNENWNLISHPDYOGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAPGEGFV 420
Db 398 APPVETTYNENWNLISHPDYOGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAPGEGFV 457
Qy 421 NVTVPARRVNLPPVAVVSPQLOELTLPLTSALIDGSDTDTEIVSYHWEINGPFIEE 480
Db 458 NVTVPARRVNLPPVAVVSPQLOELTLPLTSALIDGSDTDTEIVSYHWEINGPFIEE 517
Qy 481 KTSVDSVPLRLSLNLDGPNYSFRLTVTSDSGATNSTAALIVNNAVDPVPPVANAGNHIT 540
Db 518 KTSVDSVPLRLSLNLDGPNYSFRLTVTSDSGATNSTAALIVNNAVDPVPPVANAGNHIT 577
Qy 541 LPQNSITLNGQSSDDHQIVLYEWSLPGSGEGKHVVMQGVQTPYLHLSAMQBGDITFOLK 600
Db 578 LPQNSITLNGQSSDDHQIVLYEWSLPGSGEGKHVVMQGVQTPYLHLSAMQBGDITFOLK 637
Qy 601 VTDSRQQSTAXTVIVQPNRRPPVAVAGPKELIFPVESATLDGSSSSDDHGIVFYHW 660
Db 638 VTDSRQQSTAXTVIVQPNRRPPVAVAGPKELIFPVESATLDGSSSSDDHGIVFYHW 697
Qy 661 EHVGPASAVEMENIDKAIATVTLQVGYTHFRLTVKDDQGLSSTSTLTIVAVKKNSSPPR 720
Db 698 EHVGPASAVEMENIDKAIATVTLQVGYTHFRLTVKDDQGLSSTSTLTIVAVKKNSSPPR 757
Qy 721 ARAGRHVLVLPNNSITLDGSRSTDQRIVSVLWTRDQSPAAAGDVIDGSDHSVALQLTN 780
Db 758 ARAGRHVLVLPNNSITLDGSRSTDQRIVSVLWTRDQSPAAAGDVIDGSDHSVALQLTN 817
Qy 781 LVEGYVTFHLRTDSQASDPTATVEVQDPKESGLVELTLQVGVGLTEQRKDTLVRQ 840
Db 818 LVEGYVTFHLRTDSQASDPTATVEVQDPKESGLVELTLQVGVGLTEQRKDTLVRQ 877
Qy 841 LAVLLNVLDSIKVKIRAHSDLSITVIFYVQSRPPFKVLAARVNLHMLSEKAD 900
Db 878 LAVLLNVLDSIKVKIRAHSDLSITVIFYVQSRPPFKVLAARVNLHMLSEKAD 937
Qy 901 LFPKVLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIORYIWDGESNCWEIFYVT 960
Db 938 LFPKVLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIORYIWDGESNCWEIFYVT 997
Qy 961 VLAFTLIVLTGGFTWLCICCCCKRQKTKIRKTKYTTILDNMDEQERMELRPKYGIKHRST 1020
Db 998 VLAFTLIVLTGGFTWLCICCCCKRQKTKIRKTKYTTILDNMDEQERMELRPKYGIKHRST 1057
Qy 1021 EHNSSLMWSESEFSDQDITFSEKMERGNPKVSMNGSIRNGASFYSCKDR 1072
Db 1058 EHNSSLMWSESEFSDQDITFSEKMERGNPKVSMNGSIRNGASFYSCKDR 1109

RESULT 2
Q80U39 PRELIMINARY; PRT; 1083 AA.
AC Q80U39;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MKIAA0319 protein (Fragment).
GN Name=D130043K22Rik; Synonym=mkIAA0319;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;

```

```

RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
DR EMBL; AK122246; BAC65528.1; -.
DR MGI; MGI:3036268; DI30043K22Rik.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR00601; PKD.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00089; PKD; 5.
DR PROSITE; PS00093; PKD; 2.
FT NON TER 1
SQ SEQUENCE 1083 AA; 118201 MW; 789BBFB3A7C02CE CRC64;

Query Match 76.3%; Score 4255.5; DB 2; Length 1083;
Best Local Similarity 75.8%; Pred. No. 2.5e-200;
Matches 820; Conservative 91; Mismatches 160; Indels 11; Gaps 4;

Qy 1 MAPPTGVLSLLLVTTAGCARKOCSEGRYTSNAVISNLETTRIMRYSHTFPVVDCTAA 60
Db 3 MVSPPGVLSLLLAAMAGSSQCSGRYTSNAVISNLETTRIMRYSHTFPVVDCTAA 62
Qy 61 CCDSLSSCDLAWFEGRCVLSVCPHKECEPKMGPIRSYLTFLVRPVPQALLDYGDMM 120
Db 63 CCDDLTCDLAWFEGRCVLSVCPHKECEPKMGPIRSYLTFLVRPVPQALLDYGDMM 122
Qy 121 LNRGSPGIMGDSPEIRKDLXFLGKMWGLEEMSEYXDYDELEKDLLOPSGQKPRGSA 180
Db 123 LSRGSPSGAMGDSLEDLKDLPLFLGKMGPEETTESYDEYKDLERGLLQPSNQDPRGSA 182
Qy 181 EYTDWGLLPGSEGAFFN-SSVGDSPAVPAETQDDP-----ELHYLINESASTPAKLUPE 231
Db 183 EYDPMWLLSPNEGAFNATATGDNAAEMKIQDPTPHPLDQEQALQALNESTWSPFGHSS 242
Qy 232 RSVLLPLPTTP-SGSEVLEKEKASQLOEQSSNSSGKVLMPHSHLPASLELSSVTVKES 290
Db 243 ISSVWFPSSASPLPTEEGLEGETLQLOEQSSNSSGKVLMPHSHLPASLELSSPATTEKN 302
Qy 291 PVLVTVPSTSEHSITPTPTSAAPSESTSELPISPTTAPRTVKELTVSAGDNLIIITLDPN 350
Db 303 SNFTVTPSRKSHSTPTFTSTVLGLTTPPPPLSP-TASRTVKALAVSAGDNLVLTLPDR 361
Qy 351 EVELKAFVAPAPPVETTYNENWNLISHPDYOGEIKQGHKQTLNLSQLSVGLYVFKVTVS 410
Db 362 EAEKASVEPAPPADTTYSEWLSMHPVDFQGIQENKPTLHLSQLSVGLYAFRAVS 421
Qy 411 SENAFGEFVNVTVKPARVNLPPVAVVSPQLOELTLPLTSALIDGSDTDTEIVSYHW 470
Db 422 SENAFGEFVNVTVKPARVNLPPVAVVSPQLOELTLPLTSALIDGSDTDTEIVSYHW 481
Qy 471 EEINGPFTSEKTSVDSPVRLSLNLDGPNYSFRLTVTSDSGATNSTAALIVNNAVDPVPP 530
Db 482 EEVDGPFLEGEFPADTPIRLSLNLDGPNYSFRLTVTSDSGATNSTAALIVNNAVDPVPP 541
Qy 531 ANAGPNHTITLPONSITLNGQSSDDHQIVLYEWSLPGSGEGKHVVMQGVQTPYLHLSAM 590
Db 542 ANAGPNHTITLPONSITLNGQSSDDHQIVLYEWSLPGSGEGKHVVMQGVQTPYLHLSAM 601
Qy 591 QEGDYTFOLKVTDSRQOSTAXTVIVQPNRRPPVAVAGPKELIFPVESATLDGSSSS 650
Db 602 QEGDYTFOLKVTDSRQOSTAXTVIVQPNRRPPVAVAGPKELIFPVESATLDGSSSS 661
Qy 651 DDHGIVFYHWEHVRGSPASAVEMENIDKAIATVTLQVGYTHFRLTVKDDQGLSSTSTLIVA 710
Db 662 DDHGIVFYHWEHVRGSPASAVEMENIDKAIATVTLQVGYTHFRLTVKDDQGLSSTSTLIVA 721
Qy 711 VKKKNSSPPRAGRHVLVLPNNSITLDGSRSTDQRIVSVLWTRDQSPAAAGDVIDGSD 770
Db 722 VKKKNSSPPRAGRHVLVLPNNSITLDGSRSTDQRIVSVLWTRDQSPAAAGDVIDGSD 781

```

```
Qy 771 DHSVALQTLNVEGVYTHLRLVTDQSGASDPTATVEVQDPKSGGLVELTLQVGVGOLT 830
Db 782 DHRAALQTLNVEGVYTHLRLVTDQSGASDPTATVEVQDPKSGGLVELTLQVGVGOLT 841
Qy 831 EQRDXTLVRQALVLLNVLDSDIKQKIRAHSDLSSTVIVFYQSRPPFPKVLKAAEVAARNLH 890
Db 842 EQQKXTLVRQALVLLNVLDSVKVLKQAHTDVSTVIVFYQSRPPFPKVLKAAEVAARNLH 901
Qy 891 MRLSKEKADFLFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESSNCSVALQTLNVEGV 950
Db 902 KRLSKEKEAFLFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESSNCSVALQTLNVEGV 961
Qy 951 NCEWSIFVTVLAFLLVLTGFTWLCICCCCKRQKTKIRKTKYTTILDNDMEQERMELR 1010
Db 962 NCEWSVFVAAALAUTLLTGAWSLWCICCCRRKTKIRKTKYTTILDSWDEQERMELR 1021
Qy 1011 PKYGIKHSRSTHNSLMVSEFSDQDTIFSRERKMERGNPKVSMNGSIRNGASFYSYCSK 1070
Db 1022 PKYGIKHSRSTHNSLMVSEFSDQDTIFSRERKMERGNPKVSMNGSIRNGASFYSYCSK 1081
Qy 1071 DR 1072
Db 1082 DR 1083

RESULT 3
Q7L8R6 PRELIMINARY; PRT; 707 AA.
ID Q7L8R6
AC Q7L8R6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE D073M23.3 (KIAA0319 gene product) (Fragment).
GN Name=D073M23.3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP Thorne K.;
RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL031230; CAZ0249.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR SMART; SM00089; PKD; 4.
DR PROSITE; PS00093; PKD; 1.
FT NON TR 1
SQ SEQUENCE 707 AA; 78402 MW; FCTD2A98D778DF46 CRC64;

Query Match 65.8%; Score 3671; DB 2; Length 707;
Best Local Similarity 99.9%; Pred. No. 6.4e-172;
Matches 706; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 366 TTYNENWNLISHPTDYQGEIKQGHKQTLNLSQLSVGLVFKVTYSSENAPGEGFNVTVK 425
Db 1 TTYNENWNLISHPTDYQGEIKQGHKQTLNLSQLSVGLVFKVTYSSENAPGEGFNVTVK 60
Qy 426 PARRVNLPPVAVSPQQLPLTSLALDGSQSTDDTEIYSVHWEINGPPIEKTSDV 485
Db 61 PARRVNLPPVAVSPQQLPLTSLALDGSQSTDDTEIYSVHWEINGPPIEKTSDV 120
Qy 486 SPVLRLNLDGPNYSFRLTVTDSGATNSTTAALIVNADVPPVANAGPNTHTLPQNS 545
Db 121 SPVLRLNLDGPNYSFRLTVTDSGATNSTTAALIVNADVPPVANAGPNTHTLPQNS 180
Qy 546 ITLNGNSSDDHQIYLVLEWSLPGSGKHVVMQGVQTPYLHLSAQMGEDYTFQKLVTDSS 605
Db 181 ITLNGNSSDDHQIYLVLEWSLPGSGKHVVMQGVQTPYLHLSAQMGEDYTFQKLVTDSS 240
```

```
Qy 606 ROOSTAXVTIVOPENNRPPVAVAGDPKELI FPVESATLDGSSSDDHGIVFYHWEHVRG 665
Db 241 ROOSTAVVTIVOPENNRPPVAVAGDPKELI FPVESATLDGSSSDDHGIVFYHWEHVRG 300
Qy 666 PSAVEMENIDKATATVTGLQVGTVHFRLTAVKDOQGLSSTLTVAVKKENNSPPRARAGG 725
Db 301 PSAVEMENIDKATATVTGLQVGTVHFRLTAVKDOQGLSSTLTVAVKKENNSPPRARAGG 360
Qy 726 RHVLVLPNNSITLDGSRSTDDQRIVSYLWTRDQSPAGDIVDGSNDSHVALQTLNVEGV 785
Db 361 RHVLVLPNNSITLDGSRSTDDQRIVSYLWTRDQSPAGDIVDGSNDSHVALQTLNVEGV 420
Qy 786 YTHLRLVTDQSGASDPTATVEVQDPKSGGLVELTLQVGVGOLTQERKDTLVRQLAVLL 845
Db 421 YTHLRLVTDQSGASDPTATVEVQDPKSGGLVELTLQVGVGOLTQERKDTLVRQLAVLL 480
Qy 846 NVLDSDIKVQKIRAHSDLSSTVIVFYQSRPPFPKVLKAAEVAARNLHMLSKKADFLFKV 905
Db 481 NVLDSDIKVQKIRAHSDLSSTVIVFYQSRPPFPKVLKAAEVAARNLHMLSKKADFLFKV 540
Qy 906 LRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESSNCSVALQTLNVEGV 965
Db 541 LRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESSNCSVALQTLNVEGV 600
Qy 966 LIVLTGFTWLCICCCCKRQKTKIRKTKYTTILDNDMEQERMELRPKVGIKHSRSTHNS 1025
Db 601 LIVLTGFTWLCICCCCKRQKTKIRKTKYTTILDNDMEQERMELRPKVGIKHSRSTHNS 660
Qy 1026 LMVSEFSDQDTIFSRERKMERGNPKVSMNGSIRNGASFYSYCSKDR 1072
Db 561 LMVSEFSDQDTIFSRERKMERGNPKVSMNGSIRNGASFYSYCSKDR 707

RESULT 4
Q8KL135 PRELIMINARY; PRT; 1048 AA.
ID Q8KL135
AC Q8KL135;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Expressed sequence AU040320.
GN Name=AU040320;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FVB/N; TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242503899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Maman A., Rodriguez S., Sanchez A.,
RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=FVB/N; TISSUE=Kidney;
```





RT "RIKEN integrated sequence analysis (RISA) system-384-format

RL sequencing pipeline with 384 multicapillary sequencer."

RM Genome Res. 10:1757-1771(2000).

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Heart;

RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,

RA Fukuda S., Furuno M., Hanegaki T., Hara A., Hashizume W.,

RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,

RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,

RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,

RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,

RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,

RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,

RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,

RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK084668; BAC39244.1; -

DR MGD; MGI:2140475; AU040320.

DR InterPro; IPR003961; FN III.

DR InterPro; IPR008957; FN III-like.

DR InterPro; IPR000601; PKD.

DR SMART; SM00060; FN3; 4.

DR SMART; SM00089; PKD; 5.

DR PROSITE; PS00093; PKD; 1.

KW Hypothetical protein.

SQ SEQUENCE 1085 AA; 119329 MW; 7E61DB6001A275E4 CRC64;

Query Match 43.1%; Score 2403.5; DB 2; Length 1085;

Best Local Similarity 48.3%; Pred. No. 1.6e-109; Indels 109; Gaps 16;

Matches 514; Conservative 135; Mismatches 307;

Qy 10 SLLLVLTAGCARQCSEGRYTSNAVISPNLETRIMRVSHTFPVVDCTAACDLSLSCDL 69

Db 41 SVLWLTADSESR--CQCKTLYGALRTEGE-NHLRLLAGSLPFHACRAACRDSACHA 97

Qy 70 AWPFGRCYLVSCHKENCCEPKMGPIRSYLTFLVRPVRPAQLLDYGDMMNLGRSPSGI 129

Db 98 LWWLEGMCFQADCSKQSCQPFRTDSSNSML-----IIFQKSTQTTDD 139

Qy 130 WGDSPEDIRKDLXFLGKDWGLEE-----MSEYXDDYRELEKDLLQPSGKQEP 177

Db 140 LGLLPDEPDEPHLLRLG--WGTSWRROSLGAPLTLSPSSHQSLLRD----- 186

Qy 178 GSASYTDWGLLPSEGAFNSVGVSPAPVAPETQDDPELH-----YLNESASTPAKPLP 230

Db 187 --RQRDLNVVPTHGAMQHSKVNHEEAGALSPTSAEVRKTIIVAGSFTSNHITQTEWP 244

Qy 231 ERSVLLPLPTPTSPGSEVLEKASQLOQSSNSGKEVLMPSHSLPPASLELSVTVVEKS 290

Db 245 KNVSTHPEPS-----EHSFSPVSG----- 262

Qy 291 PVLITPSTGTHSPTPTSPASPS--TPS-ELPISPTTAPR-TVKELTVSAGNLIIT 346

Db 263 ---TPQVKSTHSPSTDAPLPVAPSYATPFOASSQSTSPAPFPVVKVLSAGKSQVIT 319

Qy 347 LPDNEVELKAFVAPAPPVETTYNVEWNLISHTPTQGEIKQCHKQTLNLSQLSVGLVYFK 406

Db 320 LPKNEVQLNALVPAEPEGETTYTDWQILTIPTDYSGEVERKHSQSLSKLTGLTEYFK 379

Qy 407 VTVSSENAFGFGEVNVTVKPARVNLPPVAVVSPQLBLTLPLTSALIDGSGQSTDDTEIV 466

Db 380 VTVDQNAHGGYVNVTVKPEPRKNRPPVAVVSPQFQISLPTSTIIDGSGQSTDDDKIV 439

Qy 467 SYHHEEINGPIEIKTSVDSVPLRLSNLDPGNYSFRLTVTSDGATNSITTAALVNNAVD 526

Db 440 QYHWEELKGLREEKISDITAILKSLVPGNVTFSLTVVDSDGATNSITTAALVNNAVD 499

Qy 527 YPPVANAGPNTITLIPONSITLNGQSDDDHQIIVLYEWSLGGSGKHVNVQGVQTPYLH 586

Db 500 YPPVANAGPQVITLIPQNSITLFGNQSTDDHGIITSYEWLSLSPSSKGVKVEVQGVTPALQ 559

Qy 587 LSAMQEGDYTFOLKYVTDSSRQSQSTAXVTIVVOPENRPPVAVAGPKDELFPVESATLDG 646

Db 560 LSAMQEGDYTYQLTVDTDAGQQAQVTVIVQVPENNKPPQADAGPKDELFPVDSITLDG 619

Qy 647 SSSDDHGIIVFYHWEHVRGSPSAVEMENIDKAIATVTGLQVGTGTVHFRILTVDKQDQGLSST 706

Db 620 SKSTDDQRVVSYLWEQSRGPDGVQLENANSSVATVGLQVGTVFTLTVDKERNLQSQSS 679

Qy 707 LTVAVKKNENSPRARAGGRHVLVFNNSITLDSRSTDDQRIVSYILWIRGQSPAGADV 766

Db 680 VNVIVKEINKPPVAKIAGNVVTLPTSTAEELDSRSSDKGIVSYLWTRDTSPPAGEV 739

Qy 767 IDGSDHSHVALQTLNVEGVTFHLRVTDSCASDPTATVEVQDPKSGLVELTLOVGV 826

Db 740 LNHSDHHPVLFLNVEGTYTFHLKVTDAKESDTRTVEVKDPPKSNLVEILLDNNV 799

Qy 827 GQLTEQRKDTLVRQLAVLNLVDSIDIKVQKIRAHSDLSITVIVFVQSRPPPKVLKAAEVA 886

Db 800 SOLTERLKGMILRIQIGVLLGLVDSIIIVKIQIPTYEQSTKMLPFVQNDPPHQLFKGEVA 859

Qy 887 RNLMRLSKKADFLPKVLRVDTAGCLLKSGHGHCDPLTKRCICSHLWNLQRIYIW 946

Db 860 AMLKSELQKQKADFLIFRALEISTVTQCLNCSDHGHCDSTFKRCVCDPFWMENFIKQLR 919

Qy 947 DGSNCEWSJFYVTVLAFTLVLGTGFTWLCICCCRQKRTKIRKTKYTITLDNMDQER 1006

Db 920 DGSNCEWSVLYVIIASFVIVAGILSWTTCRCKQK-GKPKRKYKILDATD-QES 977

Qy 1007 MEALRP--KYGIKHSRSTHNSLMVSESEFSDQDTIFSRKMERG 1049

Db 978 LELKPTSRAGSKQKPTLSLSSLMHSESELDSD-DAIFTWPDREK 1021

RESULT 6

ID Q8BHZ3 PRELIMINARY; PRT; 1085 AA.

AC Q8BHZ3;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730047D20 product:HYPOTHETICAL 103.9 kDa PROTEIN homolog.

DE Name:AU040320;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

RA Carninci P., Hayashizaki Y.,

RT "High-efficiency full-length cDNA cloning.";

RL Meth. Enzymol. 303:19-44(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;

RA RIKEN FANTOM Consortium;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Cerebellum;



RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raba S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska J., Smalut D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RT proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RL Strauberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC014530; AAH14530.1; -;  
DR SMART; SM000601; PKD;  
DR InterPro; PS00093; PKD; 1.  
DR PROSITE; PS00093; PKD; 1.  
SQ SEQUENCE 1049 AA; 115584 MW; 5BDB6057CC09A649 CRC64;  
Query Match 43.0%; Score 2402; DB 2; Length 1049;  
Best Local Similarity 48.0%; Pred. No. 1.8e-109;  
Matches 519; Conservative 134; Mismatches 313; Indels 116; Gaps 17;  
Qy 10 SLLLVTTIAGCARQCSEGRYSNAVISPNLETRIMRVSHTFPVWDCTAACCDLSCDL 69  
Db 41 SVLWLTDA--SESRCCQKGTQFVGLRSGE-NHLWLEGTPLQSLAACQDSACHV 97  
Qy 70 AWFPEGRCYLVSCPHKENCEPKMGPIRSYLTFLVRPVPQPAQLLDYGMMLNRGSPGI 129  
Db 98 FWMLEGMCIQADCSRPOSCRAFRTSHSSNSMLVP-LKKFQTADDL----- 140  
Qy 130 WGDSPEDIRKDLXFLGKDWGLEEMSEYXDDYRELEKDLLOPSGKQEPGRGSAHYTDWGLLP 189  
Db 141 -GFLPEDDDVPHLLGLGNWA-----SWRQSPRAAL----- 170  
Qy 190 GSEGAFNSSVGDSPAVPAETQDDPELHYLNESASTPAPKPER----SVLLPLPTPSSG 245  
Db 171 -----RAVSSSDQ-----SLRKQKRGSPSDVPTVPTQHS-- 204  
Qy 246 EVLEKEKASQLQEOSSNSGKVELMPHSLPPASLELS-----SVTVKSPVLTVPFGS 299  
Db 205 KYNDSNELGGITTSCEAEVHKAIITSSPLTDLTAEISGGPKNVSVQPEISEGLATTFST 264  
Qy 300 -----TEHSIFPP-----PTSAAPSESTPSELPTSPPTAPRVKELTVSAGDN 342  
Db 265 QQVKSSSEKTQIAVPQVAPSYATPTTQASFSQTSAPYPV-----IKELVVSAGES 316  
Qy 343 LIITLPDNEVELKAFVAPAPVETTYNENWILSHPTDYQCEIKQGHQTLNLQSLVGL 402  
Db 317 VOITLPKNEVQLNAYVLQEPPEKGTYYTDWQLITHPRDYSGEMEGKHSQILKJLTPGL 376  
Qy 403 YVFKVTVSSNAFGEFGFVNVTVKARRVNLPPVAVSPOLQELTLPLTSALIDGQSQSTD 462  
Db 377 YEPKIVVEGQANAGEYVNVTVKPEPRKNRPPIALVSPQFSEISLPTTSVIDGSQSTD 436  
Qy 463 TEIVSYHWEENGPIEKTSDVSPVLKSLNLDPGNYSFRLTVTDSOGATNTSTAALVN 522  
Db 437 DKIVQYHWEELKPLREKISEDTAILKSLKLVPGNYTFTSLTVVDSOGATNTSTAALVN 496  
Qy 523 NAVDPPPVANAGPNHTITLPQNSITLNGQSSDDHQIVLYEWSLGPSEGKHVWQGVOT 582  
Db 497 KAVDPPPVANAGPNQVITLPQNSITLFGNQTDDHGTISYEWLSLSPSKGKVMQGVRT 556  
Qy 583 PYLHLSAMQEGDYTFOLKVTDSRRQOSTAXVTYVQPNENRPPVAVAGPKDELFPVESA 642  
Db 557 PTLQLSAMQEGDYTYQLTVTDTIGQATAQVTVIVQPNENRPPQADAGPKDELTPVDST 616

Qy 643 TLDGSSSDDHGIVFYHWEHVRGSPSAVEMENIDKAIATVTGLQVGYHFRLTUVKDOQGLS 702  
Db 617 TLDGSSSDDHGIVFYHWEHVRGSPSAVEMENIDKAIATVTGLQVGYHFRLTUVKDOQGLS 702  
Qy 703 STSLTAVAVKKNNSPPRAGRGHVLVLPNNSTLTDGSRSTDDQRIVSVYLIWIRDGSPA 762  
Db 677 SQSSVNVVKEEINKPPIAKITGNVITLPTSTAEALDGSKSSDDKGVSVLWTRDEGSPA 736  
Qy 763 AGVIDGSDHSVALQLTNLVEGVYTFHLRVTDGSGASDTATATVEVQDDPKSGVLVELTL 822  
Db 737 AGEVLNHSDDHPLFLFLSNLVEGVYTFHLRVTDGSGASDTATATVEVQDDPKSGVLVELTL 822  
Qy 823 QVGVQLTQTEQRKDTLVRQLAVLLNLDSDIKVQIRAHSDLSVTIVFVQSRPPPKVLKA 882  
Db 797 DINVQLTERLKGMPFIRQIGVLLGLVDSIIIVQIPTYTEQSTKGVFVQNEPQHIKFG 856  
Qy 883 AEVARNLHMLRSKEKADFLFKVLVRVDTAGCLLKCSGHGCDPLTKRCICSHLWMLNIQ 942  
Db 857 HEVAAMLKSELKQKADFLIFRALEVNVTVCQLNCSDHGCHDSFTKRCICDPFWMENFIK 916  
Qy 943 RYIWDGSDNCESWIFVYTVLAFTLIVLTGGFTWLCICCKRQKTKIRKTKYTYILDND 1002  
Db 917 VQLRDGSDNCESWVLYYIATFVIVVALGILSVTWVICCKRQK-GKPKRSKYKILDATD 975  
Qy 1003 EQRMEELRP--KVGIKHRSTHNSSLMVSSSEPSDDDTIFSRKEMRGNPKVSMNGSIR 1060  
Db 976 -QESLELKPSTRAGIKQKGLLSLHSSSELDSD-DAIFTPWDRKGLLHGQNGSVP 1033  
Qy 1061 NG 1062  
Db 1034 NG 1035  
RESULT 8  
Qy Q81ZAO PRELIMINARY; PRT; 1049 AA.  
Db Q81ZAO  
AC Q81ZAO;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DB Polycystic kidney disease 1-related protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
ON NCBI\_TaxID=9606;  
RX [1]  
RP SEQUENCE FROM N.A.  
RL Hao D., Hooi S.;  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY163234; AAN61054.1; -;  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR000601; PKD.  
DR SMART; SM00060; FN3; 4.  
DR SMART; SM00089; PKD; 5.  
DR PROSITE; PS00093; PKD; 1.  
SQ SEQUENCE 1049 AA; 115717 MW; 1781BFCA3C765DEB CRC64;  
Query Match 43.0%; Score 2401; DB 2; Length 1049;  
Best Local Similarity 48.0%; Pred. No. 2e-109;  
Matches 519; Conservative 134; Mismatches 313; Indels 116; Gaps 17;  
Qy 10 SLLLVTTIAGCARQCSEGRYSNAVISPNLETRIMRVSHTFPVWDCTAACCDLSCDL 69  
Db 41 SVLWLTDA--SESRCCQKGTQFVGLRSGE-NHLWLEGTPLQSLAACQDSACHV 97  
Qy 70 AWFPEGRCYLVSCPHKENCEPKMGPIRSYLTFLVRPVPQPAQLLDYGMMLNRGSPGI 129  
Db 98 FWMLEGMCIQADCSRPOSCRAFRTSHSSNSMLVP-LKKFQTADDL----- 140  
Qy 130 WGDSPEDIRKDLXFLGKDWGLEEMSEYXDDYRELEKDLLOPSGKQEPGRGSAHYTDWGLLP 189  
Db 141 -GFLPEDDDVPHLLGLGNWA-----SWRQSPRAAL----- 170



```
Db 755 EVAAMKSELKQKADFLIFRALEVNTVTCQLNCSHGHCDSFTKRCICLDPFWMFNFKV 814
Qy 944 YWPGSCNWSIFVYVTLVLAFTLVLTGFTWLCICCCCKRQRTKIRKTKYITILDNMDE 1003
Db 815 QLRDGSNCNSVLVILATFVIVVALGILSWTVICCCCKRQK-GKPKRKSXYKILDATD- 872
Qy 1004 QERMELRP--KYGKIRHSTHNSSLMVSESEFSDSDODTIFSRXKMERGNPKVSMNGSIRN 1061
Db 873 QESLEKPTSRAGIKQKGLLSLSSLMHSESELDSD-DAIFTWPDREKGLLHGQNGSVFN 931
Qy 1062 G 1062
Db 932 G 932

RESULT 10
QNDAAO
ID Q8NDAO PRELIMINARY; PRT; 746 AA.
AC Q8NDAO;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein DKFZp434C0829 (fragment).
GN Name=DKFZp434C0829;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RG The German cDNA Consortium;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834315; CAD38985.1; -
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR00601; PKD.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00089; PKD; 5.
DR PROSITE; PS50093; PKD; 1.
KW Hypothetical protein.
SQ NON TER 1
SQ SEQUENCE 746 AA; 82461 MW; A0945E57D8243191 CRC64;

Query Match 40.4%; Score 2256.5; DB 2; Length 746;
Best Local Similarity 60.2%; Pred. No. 1.6e-102;
Matches 441; Conservative 98; Mismatches 189; Indels 5; Gaps 4;

Qy 332 VKELTVSAGDNLIIITLPDNEVELKAFVAPAPPVETTYNNYENWLISHPTDYQGEIKQGHKQ 391
Db 3 IKELVVSAGESVQITLPKNEVQLNAYLQEPKGETTYTDYDQLTHPRDYGEMEGKHSQ 62
Qy 392 TLNLQSLVGLYFVKVTVSSNAFGEFVNVTVPARRVNLPPVAVSPQQLTLPPTS 451
Db 63 ILKLSKLPGLYFVKVIVEGQNAHGEVYVTVKPEPRKNRPPIAIVSPQQLTLPPTS 122
Qy 452 ALIDGSQSDTDEIVSYNWEINGPFIEKTSVDSVPLRLNLDPCNYSPLTIVTDSGA 511
Db 123 TVIDGSQSDTDDKIYQVHWEELKGPLREEKISDAILKLSKLVPGNYSPLTIVTDSGA 182
Qy 512 TNSTTAALIVNNAVDPYPPVANAGPNHITLTPNSITLNGNOSDDHQVLVYEWLSLPGSE 571
Db 193 TNSTTANLTVKAVDPPVANAGPNQVITLPQNSITLFGNSTDHDGHTSTWLSLPSGK 242
Qy 572 GKHVVMQGVQTPYLLHLSAMQEGDYTFOLKVTDSRQSQSTAXVTIVQPNRPPVAVAGP 631
Db 243 GKVEMQGVRTPTLQLSAMQEGDYTYQLTVDITIGQATAQVTVIVQPNRPPQADAGP 302
Qy 632 DKELIFPVESATLDGSSSDHGIYFVYHWEHVRGPSAVEMENIDKAITVTLQVGTTHF 691
```

```
Db 303 DKELTLPVDSITLTDGSKSSDDQKIISYLWEKTQCPDGVQLENANSSVATVTLQVGTVP 362
Qy 692 RLTVDDQOGLSSSTLTVAVKKNNSPPRAGRGHVLVLPNNSITLDGSRSTDDQRIYS 751
Db 363 TLTVDKBERNLQSSSVNVIVKEINKPPIAKITGNVITLPTSTAELDGSKSSDDKGIVS 422
Qy 752 YLWTRDGSQSPAAGDIVDGSVHALQLTNLVGEVYVTFHLRVTDQSGASDITDTATVEQPD 811
Db 423 YLWTRDGSQSPAAGEVNLHSDHHPILFNLVGEYTYFHLKVTDAKGSDDTDTTVEVKPD 482
Qy 812 PRKSLVELTLQVGVGQLTQEQKDTLVRQLAVLNLVDSDIKVKQIRAHSDLSVIVFY 871
Db 483 PRKNLVEIILDINVSQLTERRLKGWIFRQIGVLLGLVLDSDIIVQIKIPTYEQSTKMWFFV 542
Qy 872 QSRPPFKVLKAAEVARNLHRLSKKADFLFKVLRVDTAGCLLKCSHGHCDDPLTKRCI 931
Db 543 QNEPPHQIFKGEHEVAAMLKSELKQKADFLIFRALEVNTVTCQLNCSHGHCDSFTKRCI 602
Qy 932 CSHLWMENLIQRYIWDGSCNWSIFVYVTLVLAFTLVLTGFTWLCICCCCKRQRTKIRK 991
Db 603 CDPFWMFNFKVQRDGSNCNSVLVILATFVIVVALGILSWTVICCCCKRQK-GKPKR 661
Qy 992 KTKYTILDNMDEQERMELRP--KYGKIRHSTHNSSLMVSESEFSDSDODTIFSRXKMERG 1049
Db 662 KSKYKILDATD-QESLEKPTSRAGIKQKGLLSLSSLMHSESELDSD-DAIFTWPDREK 719
Qy 1050 NPKVSMNGSIRNG 1062
Db 720 KLLHGQNGSVPNG 732

RESULT 11
Q9H7V0
ID Q9H7V0 PRELIMINARY; PRT; 691 AA.
AC Q9H7V0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein FLJ14225 (PKD1-like protein).
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiroaka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Mueshino K., Yuuki H., Oshima A., Sasaki N., Aoteuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
```

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RL Nat. Genet. 36:40-45(2004).

SEQUENCE FROM N.A.

RC TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahney J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley A.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

SEQUENCE FROM N.A.

RC TISSUE=Kidney;  
RA Strausberg R.;  
RL Submitted (JUN-2002) to the ENBL/GenBank/DBSJ databases.  
DR ENBL; AK024287; BAB14874.1; -  
DR ENBL; BC031672; AAB31672.1; -  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR00601; PKD.  
DR SMART; SM00060; FN3 3.  
DR SMART; SM00089; PKD 4.  
DR PROSITE; PS00093; PKD 1.  
DR PROSITE; PS00093; PKD 1.  
SQ SEQUENCE 691 AA; 76234 MW; E3B9768BEE20F39 CRC64;

Query Match 37.4%: Score 2086.5: DB 2: Lenath 691:

Best Local Similarity 60.7%; Pred. NO. 3.1e-94; Matches 410; Conservative 88; Mismatches 173; Indels 5; Gaps 4;

Qy	389	HKQTNLNLSQLSVGLYFKVTKTVSS	ENAFGEGFVNVTVKPARRVNLPPVA	VVVSPOQL	ETLP	448
Db	5	HSQILKLSKLTGPIYEFKVI	VEGNAHGEYVNVTVKPEPKNRPIA	IVSPOFOE	ISLP	64

Qy	449	LTSALIDGQS	TDDTEIVSYHWEENGPFIEEKTSVDSPVLRLSNLDPNGYSFRLTVTDS	508
Dp	65	TTSTVINGDS	TDDKIVYHWBELKGPIREEKISEDATLKLSKLPNGYTSLTVTDS	124

Qy 509 DGATNSTTAALIVNNAVDYPPVANAGPNHITITLPQNSITLNGNSSDDHQIVLYEWSLGP 568

Dh 125 DGATNSTTANI,TVNKAVDYPPVANAGPNOVITITLPQNSITIT,FGNSTDDHGI,TSVEWSI,SP 184

Qy	569	GSEGHVVHVGQGVQTPYHL	SAMQEGDYTFOLKVTDSRRQOSTAXVTIVIQPENRPPAV	628
Dh	195	SSSGKVVHVGQGVQTPYHL	SAMQEGDYTFVLTFTICQSTAXVTIVIQPENRPPORD	244

QY 629 AGPKELIPVESATLDGSSSSDHGIVFYHWEHVRGSPSAVEMENIDKAIATVTLQVGT 688

nb 245 AGPDVEIETPVDSSTIDGSSSSDHGIVISVWEVTCGPRGVIVFVANSVATVTLQVGT 304

Qy	YHFLTVKDOOGLSSTLTVAVKENNSPPRAGGRHVLVPNNSITLDGSRSTDDOR	748
689	YHFLTVKDOOGLSSTLTVAVKENNSPPRAGGRHVLVPNNSITLDGSRSTDDOR	748
205	YVETITVTDDEBNICSSGVNVVETFEINFDIVTCTVNVITITETCTATCTGCGGCGDVC	204
204	YVETITVTDDEBNICSSGVNVVETFEINFDIVTCTVNVITITETCTATCTGCGGCGDVC	204

QY 749 IVSYLWIRDGQSPAAGVIDGSDHSHVALQLTNLVEGVYTFHLRVTDSCGSDTDTATVEV 808

365	Db	IVSYLWTRDEGSPAAGEVLNHSDDHPILFLSNLVEGTIFYPHLKVTDAKCGESDTRTIVEV	424
809	Qy	QDPDRKSGVLVELTLQVGVGQLTEQRKDTVLVROLVALLNLVSDSIKVQKIRAHSDLSSTVIV	868
425	Db	KPDPKRNVLVEITLDINVSQLTERLKGMPIROIGVLVLGVLDSDIIVQKIQPYTQEOSTKMW	484
869	Qy	FYVQSPPPPFKVLKAAEVAARNLHWRLSKKADFLFLKVLRVDTAGCLLKCSGHGCHCDPLTK	928
485	Db	FFVQNEPPHQIPKGEVAAVLKSELKQKADFLIFRALEWNTVTCQLNCSDHGCHDSFTK	544
929	Qy	RTCSSHLMNENLQRYIWGBSNCWSIPYVTVLATLVLTCGFTWLVICCCCKROKRTK	988
545	Db	RCICDPFMWENFIKVLQRDGDNCWSVLVYIATFVIVVALGILSWTVICCCCKROK-GK	603
989	Qy	IRKTKYTYLLDNWDEOERMELRP--KYGIKHRSTENHSLMWSESFSDSDODDTIFSRKEM	1046
604	Db	PKRKSXYKILDATD-QESLELEKFTSRAGIKQKGLLLSSSLMHSESELDSD-DAIFTWPDPR	661
1047	Qy	ERNPKVSMNGSIRNG	1062
662	Db	EKGKLLHGONGSPVNG	677

## RESULT 12

K96JJI		PRT; 639 AA.
ID	Q96JJU	PRELIMINARY;
AC	Q96JJU;	
DT	01-DEC-2001	(TREMBlrel. 19, Created)
DT	01-DEC-2001	(TREMBlrel. 19, Last sequence update)
DT	01-MAR-2004	(TREMBlrel. 26, Last annotation update)
DE	KIAA1837 protein (Fragment).	
GN	Name=KIAA1837;	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Theria; Primates; Catarrhini; Hominoidea; Homo.	
OX	NCEI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	

RC ISSUE=BLANK;  
PY MEDLINE=21245130.

RA	MSLINES=21245350; PUBMED=11347506;	
RA	Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;	
RT	"Prediction of the coding sequences of unidentified human genes. XX.	
RT	The complete sequences of 100 new cDNA clones from brain which code	
RT	for large proteins in vitro."	
RL	DNA Res. 8:85-95(2001).	
DR	EMBL: AB058740; BAB47466.1; -.	
DR	InterPro: IPR003961; FN III.	
DR	InterPro: IPR008957; FN III-like.	
DR	InterPro: IPR00601; PKD.	
DR	SMART: SM00060; FN3; 3.	
DR	SMART: SM00089; PKD; 4.	
DR	PROSITE: PSS0093; PKD; 1.	
FT	NON TER	1
SO	SEQUENCE	639 AA: 70427 MW: 0BE8348687170017 CRC64:

Query Match

Query Match  
Best Local Similarity  
Best Local Similarity  
Score 2004, BB 2, Length 000  
55.5%,  
62.0%: pred. No. 3e-90:

Best local similarity 92.0%, Area: NO: 35-50,  
Matches 386: Conservative 82: Mismatches 153: Indels

QY 389 HKQTNLISQLSGLYVFKVTSSENAFGEFVNVTVKPARRVNLPPVAWSQLOELTLP 448

dh 5 HSGTILKTSKLTGELYFKVIVGEGONHAGGEGVNVTVKPEPPKPPBPTATVSPFOFRTSLP 65

QY	449	LTSALIDGSQSDTDDTEIVSYHWEIINGPFFIEKTSVDSPLVLRSLNLDPGNYSFRLTVTDS	508
pb	56	TSSTVIVIDGSSQSDTDDTEIVSYHWEIINGPFFIEKTSVDSPLVLRSLNLDPGNYSFRLTVTDS	125

[illegible]

QY 569 GSEGHVVMGVQTPYLHLSAQEGDYTFQLKVTDSRQOSTAXVTVIIVPENNRPPVAV 628



Db 186 SSKGVVEMQVTRPTLQLSAMQSGDYTYQLTVTDTTIGQQATAQVTVIVPENNNKPPQAD 245  
Qy 629 AGDPKELIFPVESATLDGSSDDHGIYFVHWEHVRGSPASAVEMENIDKAIATVTVGLQVGT 688  
Db 246 AGDPKELTLPVDSTTLTGSKSDDKIISYLWERTQDGPVQLENANSSVATVTVGLQVGT 305  
Qy 689 YHFLRTVKDQGLSTSTLTAVKKNNSPPRAGRGHRLVLPNNSTLTLGSRSTDDOR 748  
Db 306 YVFTLTVDERNLQSSQSVNVIVKEEINKPPIAKITGNVITLTSTAEALDGSXSDDKG 365  
Qy 749 IVSYLWIRIDGQSPAGDVIDGSDHSAVALQTLNVEGVYTFHLRVTDGSGADTDATVTV 808  
Db 366 IVSYLWIRIDGQSPAGDVINLSDHHPILFLSNLVEGVYTFHLRVTDGSGADTDRTTVEV 425  
Qy 809 QPDRKSGELVELTLQVGVQGLTEQRKOTLVRQLAVLNLVDSIDKVKQIRAHSDLTSTIV 868  
Db 426 KPDPRKNLVEIILIDINVSQLTERLKGFIQIGVLLGLVSDIIVQKIQPTVQSTQMW 485  
Qy 869 FYQSRPFPKVLKAAEAVARNLHMLSKKADFLPKVLRVDTAGCLLKCSGHGCHDPLTK 928  
Db 486 FFVQNEPPOHIFKGHEVAAMLKSELKQKADFLIFRALEVNVTVCQLNCSDHGCHDSFTK 545  
Qy 929 RCICSHLWENLIQRYIWDGESNCESWISFYVTVLAFTLIVLTGFTWLCICCCCKRQRTK 988  
Db 546 RCICDPFWMFNIKVLQDRGDSNCESWLVYIATFVIVVALGILSVTIVCCCKRQK-GK 604  
Qy 989 IRKTKYKTLIDNMQERMLRP 1011  
Db 605 PKRSKYKILDATD-QESLELKP 626

RESULT 13  
Q8VBZ9 ID Q8VBZ9 PRELIMINARY; PRT; 603 AA.  
AC Q8VBZ9;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE AU040320 protein (Fragment).  
GN Name=AU040320;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=FVB/N; TISSUE=Liver;  
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Frange C.,  
RA Raha S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blackley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grinchwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Liver;  
RA Strausberg R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC022154; AAH22154.1; -.  
DR MGD; GI:2140475; AU040320.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.  
DR SMART; SM00060; FN3; 2.  
DR SMART; SM00089; PKD; 3.  
DR PROSITE; PS00093; PKD; 1.  
FT NON\_TER 1  
SQ SEQUENCE 603 AA; 65967 MW; BE0DA1665D4046BD CRC64;  
  
Query Match 30.0%; Score 1675.5; DB 2; Length 603;  
Best Local Similarity 59.4%; Pred. No. 3.6e-74;  
Matches 328; Conservative 76; Mismatches 143; Indels 5; Gaps 4;  
  
Qy 513 NSTTAALIVNADVPPVAVAGNPHHTITLPNSITLNGNOSDDHDIQVLYEWSLGPSEGG 572  
Db 41 NSTTASLTAVKADVPPVAVAGNPHHTITLPNSITLFGNQSTDDHIGITSYEWLSLSPSSKG 100  
Qy 573 KHVMQGVQTPYLHLSAMQSGDYTFOLKVTDSRQOSTAXVTVIVOBENRPPVAVAGPD 632  
Db 101 KVVEMQGVTRPALQLSAMQSGDYTYQLTVDTAGQATAQVTVIVQENNNKPPQADAGPD 160  
Qy 633 KELIFPVESATLDGSSDDHGIYFVHWEHVRGSPASAVEMENIDKAIATVTVGLQVTVHFR 692  
Db 161 KELTLPVDSITLDGSKSTDQRRVSYLWESQSGDPGVQLENANSSVATVTVGLQVTVYFT 220  
Qy 693 LTVKDDQGLSTSTLTAVKKNNSPPRAGRGHRLVLPNNSTLTLGSRSTDDORIVSY 752  
Db 221 LTVKDERNLQSSQSVNVIVKEEINKPPIAKITGNVITLTSTAEALDGSRSDDKGIYSY 280  
Qy 753 LWRDQSGPAAGDVIDGSDHSAVALQTLNVEGVYTFHLRVTDGSGADTDATVTVQVQDP 812  
Db 281 LWRDETSPAAGEVLNLSDDHHPVLFSLNVEGVYTFHLRVTDGSGADTDRTTVEVRKDDP 340  
Qy 813 RKSGLVELTLQVGVQGLTEQRKOTLVRQLAVLNLVDSIDKVKQIRAHSDLTSTIVFVQ 872  
Db 341 RKSNLVEIILDVNSQUTERLKGMLIRQIGVLLGLVSDIIVQKIQPTVQSTQMWFFVQ 400  
Qy 873 SRPFPKVLKAAEAVARNLHMLSKKADFLPKVLRVDTAGCLLKCSGHGCHDPLTKRCIC 932  
Db 401 NDPPHQLFKGHEVAAMLKSELKQKADFLIFRALEISTVTCQLNCSDHGCHDSFTKRCVC 460  
Qy 933 SHLWENLIQRYIWDGESNCESWISFYVTVLAFTLIVLTGFTWLCICCCCKRQRTKIRK 992  
Db 461 DPFWMENFIKVLQDRGDSNCESWLVYIATFVIVVALGILSVTIVCCCKRQK-GKPKRK 519  
Qy 993 TKYTLIDNMQERMLRP--KYGIKHSSTHNSLSLVSESEFSDSDDTIFSRKMERGN 1050  
Db 520 SRYKILDATD-QESLELKPTRAGSKOKGPTLSLSLHSESELDSD-DATFTWPDREKQK 577  
Qy 1051 PKVSMNGSIRNG 1062  
Db 578 LLYGQSGSVENG 589  
  
RESULT 14  
Q8BHR5 ID Q8BHR5 PRELIMINARY; PRT; 684 AA.  
AC Q8BHR5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Mus musculus 7 days embryo whole body cDNA, RIKEN full-length enriched  
DE library, clone:C430042D23 product:HYPOTHETICAL 103.9 kDa PROTEIN  
DE homolog (Fragment).  
GN Name=AU040320;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.



RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Franckoch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borikova D., Botchan M.R., Bouch J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of *Drosophila melanogaster*."  
 RA Science 287:2185-2195(2000).  
 RA [2]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RA "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*  
 RT melanogaster euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RN MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RN a genomic perspective."  
 RN Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celinker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RN SEQUENCE FROM N.A.  
 RP FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.

RG FlyBase;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AEO03556; AAF50494.1; -  
 DR FlyBase; FBgn0035833; CG7565.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR00601; PKD.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00089; PKD; 5.  
 DR PROSITE; PS01186; EGF 2; UNKNOWN 1.  
 SQ SEQUENCE 1069 AA; 117310 NW; 2BA50A62EE463DB CRC64;  
  
 Query Match 26.08; Score 1451.5; DB 2; Length 1069;  
 Best Local Similarity 32.1%; Pred. No. 7.5e-63;  
 Matches 373; Conservative 172; Mismatches 233; Gaps 35;  
  
 QY 8 LSSLLLVLTAGCA-----RKCQSEGRYTSNAVISNLETTRIMRVSHST 51  
 DB 8 ICNLLLTATAMSSAYADVTTQNALLVGSKKHETSPDNSVGGSS-ISPMLVCHKMLR--HV 64  
 QY 52 F-----PVVD-----CTAACCD-----LSSCDLAWWPEG 75  
 DB 65 FENATPRDEQOAGVFEBYKPPDAVEPLEBEAYLWNCLOACCEKPRNGSSACNVLVFKA 124  
 QY 76 RCVLVSCHPKENCEPK---KMGPIRSYLTFLVRPV-----QRPQLLDYGD 118  
 DB 125 KCHIRQCOSNEACLPKLRVRM-PNEKQVMVLNPLGDATWPDLLKAEAKQNALIPYDE 183  
 QY 119 MMLNRGSPGSIWGDSPRIDRKDLXFLGKOWGLEMSYXDD-----YRELEKDLLQPSGK 173  
 DB 184 AALN-----FWKQP-----RLSYLARN---QETPVYEDDFPLADKRWQMTFQPDEN 229  
 QY 174 QEPGRSABYTDWGLLPGESEGFNSVGDSPAPVAPBATOQDPDELHYLNBSA-----STPA 226  
 DB 230 -----DVLANEELGYDSSNAKFTTCDMETPC 255  
 QY 227 PKLPERSVLPLPTPSS-----GEVLEKEK-----ASOLEQSSSSSGKEV 268  
 DB 256 PP-PQCCV-----PLQPNAGVCTCPGFWNNKORKCMAAVPYSSYL---TSNAGQOE 307  
 QY 269 LMPSHSLPPASLESLSVTVEKSPVLVTPGSTHSTSIPTPTSAAPSESTPSELSPITTA 328  
 DB 308 AAASENSPEVS-----TPPLKAEQN----- 327  
 QY 329 PRTVKELTVSAGDNLIITLPDNEVELKAFVAPAPPV-ETTYNVEWNLISHT-DYQGEIK 386  
 DB 328 ---KDIWVSWSK-EVRLPEQEVTLAAFTVPDQSDTKYKYLWTLISQKPGMNGTIS 382  
 QY 387 QGHQKTNLSQLSVGLYVFKVTYSSEN-AGEGFWNTVVKPARVNLPPVAVVSPQLQEL 445  
 DB 383 DQSKSKVKLSNLSEGLYTFKVTVDGNGTGEATANTVLPENRINQPPQVILSPREQII 442  
 QY 446 TLPLTSALIDGSOSTDTEIVSVHWEISINGPFTEKTSVDSPLRLSLNDPGNYSFRLTV 505  
 DB 443 RQPTTNAIDGSTSTDDKITNHWHEVIGPIGVQVLPVENVTLQDLTSPENTFKLTV 502  
 QY 506 TSDSGATNSTTAALIVNNAVVDYPPVANAGPNHTITLPQNSITLNGNSSDDHQHVLVYEW 565  
 DB 503 TDSNNVNTSTATTIAVLKETDYAPVANAGDAVILYLPNNVNTLNGTASSDDHVEIAMEWT 562  
 QY 566 LGPGSEGHVVMQVQTPYHLHLSAQMGDGYFQLKVTDSRQOSTAXVTIVQENRPP 625  
 DB 563 KDAASDEAKAVDMQNTRTFYVQLSNLEBGMTFVLKVTGDSGQSSTAKVHVFPKPTNSPP 622  
 QY 626 VAVAGDPKELIFPVESATLDGSSSSDDHGLVFWHWEHVRGPSAVEMENIDKAIATVGLQ 685  
 DB 623 VAERAGNTTSLPINWLLNGSSDKDGIKSYLWKQLSGPNNAVILKSSNSIANATSLT 682  
 QY 686 VGTYHFLTVKDCQGLSSTLTAVVAKENNSPPRARGRHVLVLPNNSTLTDGSRSTD 745  
 DB 683 LGLYEFELTVADENNNTATDTTWKVIQVORNAAPIANAGGDHTVTLPTATYIFNGSKSWD 742  
 QY 746 DQRIVSYLWTRDQGSPPAAGDVLDGSDHSVALQLTNLVEGVYTFHLRVTDSOGASDSTDTAT 805



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:00:24 ; Search time 908 Seconds

(without alignments)  
491.154 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSLLLVITAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1854112 seqs, 416015017 residues

Total number of hits satisfying chosen parameters: 1854112

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US11A\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5576	99.9	1072	16	US-10-408-765A-1622
2	5576	99.9	1072	16	US-10-764-390-11
3	5574	99.9	1072	16	US-10-764-390-3
4	5574	99.9	1072	16	US-10-764-390-8
5	5574	99.9	1072	16	US-10-764-390-9
6	5574	99.9	1072	16	US-10-764-390-12
7	5574	99.9	1072	16	US-10-764-390-259
8	5574	99.9	1072	16	US-10-764-390-273
9	5485	98.3	1053	16	US-10-764-390-273
10	5485	98.3	1053	16	US-10-764-390-274
11	5485	98.3	1063	16	US-10-764-390-7
12	5485	98.3	1063	16	US-10-408-765A-1622
13	5485	98.3	1063	16	US-10-764-390-11
14	5485	98.3	1063	16	US-10-764-390-3
15	5485	98.3	1063	16	US-10-764-390-8
16	5485	98.3	1063	16	US-10-764-390-9
17	5485	98.3	1063	16	US-10-764-390-12
18	5485	98.3	1063	16	US-10-764-390-259
19	5485	98.3	1063	16	US-10-764-390-273
20	5485	98.3	1063	16	US-10-764-390-274
21	5485	98.3	1063	16	US-10-764-390-7

12	5485	98.3	1063	16	US-10-764-390-10	Sequence 10, Appl
13	5485	98.3	1063	16	US-10-764-390-272	Sequence 272, App
14	1451.5	26.0	1069	20	US-11-097-143-16884	Sequence 16884, A
15	1375	24.6	711	18	US-10-450-763-39353	Sequence 39353, A
16	272	4.9	155	14	US-10-050-704-175	Sequence 175, App
17	272	4.9	155	16	US-10-798-512-175	Sequence 175, App
18	232	4.2	1204	15	US-10-282-122A-49627	Sequence 49627, A
19	225	4.0	27118	17	US-10-805-684-110	Sequence 110, App
20	224	4.0	26926	9	US-09-759-508B-2	Sequence 2, Appli
21	224	4.0	26926	16	US-10-723-860-133	Sequence 133, App
22	224	4.0	26926	16	US-10-656-873A-2	Sequence 2, Appli
23	224	4.0	26926	18	US-10-756-149-4768	Sequence 4768, Ap
24	219.5	3.9	792	17	US-10-875-518-11	Sequence 11, Appl
25	216	3.9	2402	16	US-10-661-809-20	Sequence 20, Appl
26	214	3.8	36946	18	US-10-840-512-155	Sequence 155, App
27	213	3.8	7107	20	US-11-097-143-1234	Sequence 1234, Ap
28	209.5	3.8	5559	15	US-10-282-122A-75177	Sequence 75177, A
29	199	3.6	2768	20	US-11-097-143-31983	Sequence 31983, A
30	199	3.6	5701	9	US-09-864-761-37319	Sequence 37319, A
31	199	3.6	5701	14	US-10-029-386-32438	Sequence 32438, A
32	197	3.5	361	10	US-09-989-442-126	Sequence 126, App
33	197	3.5	447	17	US-10-489-695-21	Sequence 21, Appl
34	197	3.5	468	16	US-10-470-390A-18	Sequence 18, Appl
35	197	3.5	468	17	US-10-489-695-22	Sequence 22, Appl
36	197	3.5	500	15	US-10-203-875-4	Sequence 4, Appli
37	197	3.5	500	15	US-10-297-639-5	Sequence 5, Appli
38	197	3.5	550	15	US-10-369-493-10336	Sequence 10336, A
39	195	3.5	888	18	US-10-450-763-44262	Sequence 44262, A
40	195	3.5	888	18	US-10-450-763-54269	Sequence 54269, A
41	191.5	3.4	2803	15	US-10-415-187-5	Sequence 5, Appli
42	191	3.4	1194	15	US-10-282-122A-46163	Sequence 46163, A
43	188.5	3.4	1194	15	US-10-282-122A-46577	Sequence 46577, A
44	188	3.4	528	10	US-09-840-746-20	Sequence 20, Appl
45	187.5	3.4	3507	15	US-10-369-493-5784	Sequence 5784, Ap

#### ALIGNMENTS

#### RESULT 1

US-10-408-765A-1622  
; Sequence 1622, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; FILE REFERENCE: 660089.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1622  
; LENGTH: 1072  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-408-765A-1622

Query Match 99.9%; Score 5576; DB 16; Length 1072;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 MAPPTGVLSLLLVITAGCARKQCSGRYSNAVISPNLETTTRIMRVSHFTFPVVDCTAA 60  
Db 1 MAPPTGVLSLLLVITAGCARKQCSGRYSNAVISPNLETTTRIMRVSHFTFPVVDCTAA 60  
Qy 61 CCDSLCDLAWFEGRCYLVSCPHKENCEPKMGPIRSYLTFLVRPVQRPQALLDYGDM 120

Db	61	CCDSSCDLAWFFGRCYLVSCPHKENCEPKMGPIRYSYLTFLURPVQRPQAQLLDYDGM	120
Qy	121	LNRGSPGIGWGDSPEDIRKDLXFLGDKWGLEEMSEYXDDYREKOLLQPSGKQEPGRGA	180
Db	121	LNRGSPGIGWGDSPEDIRKDLTFLGDKWGLEEMSEYSDDYREKOLLQPSGKQEPGRGA	180
Qy	181	EYTDWGLLLPGSEGFANFSGVSDSPAVPAETQODPELHYLNESASTPAPKLPERSVLLPLPT	240
Db	181	EYTDWGLLLPGSEGFANFSGVSDSPAVPAETQODPELHYLNESASTPAPKLPERSVLLPLPT	240
Qy	241	TPSGEVLEKEKASQLOEQSSNSGKEVLPWSPHSLPPASILELSVTVEKSPVLTVTPGST	300
Db	241	TPSGEVLEKEKASQLOEQSSNSGKEVLPWSPHSLPPASILELSVTVEKSPVLTVTPGST	300
Qy	301	EHSIPTPTTSAAPSESTPSELPIPTPTAPRTVKELTVSAGDNLLIITLPDNEVELKAFVAP	360
Db	301	EHSIPTPTTSAAPSESTPSELPIPTPTAPRTVKELTVSAGDNLLIITLPDNEVELKAFVAP	360
Qy	361	APPVETTYNYEWNLLSHPTDQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSENAFGGFV	420
Db	361	APPVETTYNYEWNLLSHPTDQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSENAFGGFV	420
Qy	421	NVTVKPARRVNLPPVAVVVSQLOELTPLTSALIDGQSQTDDTEIVSHWEEINGPPIEE	480
Db	421	NVTVKPARRVNLPPVAVVVSQLOELTPLTSALIDGQSQTDDTEIVSHWEEINGPPIEE	480
Qy	481	KTSVDSVPVLRSLNDPGNYSFRLTVTDSGATNSTTAALIVNNAVDPYPVANAGPNHTIT	540
Db	481	KTSVDSVPVLRSLNDPGNYSFRLTVTDSGATNSTTAALIVNNAVDPYPVANAGPNHTIT	540
Qy	541	LPQNSITILNGQSSDDHQIVLYEWSLPGSGEGKHVMQGVQTPYLHLASMQEGDYTEQLK	600
Db	541	LPQNSITILNGQSSDDHQIVLYEWSLPGSGEGKHVMQGVQTPYLHLASMQEGDYTEQLK	600
Qy	601	VTDSSRQOSTAXTVTVIQQENRPPVAVAGDPKELIIPVRESATLDGSSSDDHGIVFYHW	660
Db	601	VTDSSRQOSTAVTVTVIQQENRPPVAVAGDPKELIIPVRESATLDGSSSDDHGIVFYHW	660
Qy	661	EHVRGSPASAVENIDKAIATVTGIVQGYTHFRLTVKQOGLSSTSTLTIVAKKENNSPPR	720
Db	661	EHVRGSPASAVENIDKAIATVTGIVQGYTHFRLTVKQOGLSSTSTLTIVAKKENNSPPR	720
Qy	721	ARAGRHVLVLPNNSITLDGSRSTDQRIVSYLVIRDGQSPAAGDVLDGSDHSVALQLTN	780
Db	721	ARAGRHVLVLPNNSITLDGSRSTDQRIVSYLVIRDGQSPAAGDVLDGSDHSVALQLTN	780
Qy	781	LVEGVYTHLRAVTDQSGASDdTATVEVQPPRKSGLVELTLQVGVQGLTEQRKDTLVRQ	840
Db	781	LVEGVYTHLRAVTDQSGASDdTATVEVQPPRKSGLVELTLQVGVQGLTEQRKDTLVRQ	840
Qy	841	LAVLLNVLDSDIKVQIRAHSDLSVTIVFYVQSRPPPKVLKAAEVARNLHMLRSKEKADF	900
Db	841	LAVLLNVLDSDIKVQIRAHSDLSVTIVFYVQSRPPPKVLKAAEVARNLHMLRSKEKADF	900
Qy	901	LLFKVLRVDTAGCLLKCSGHGCDPLTKRCITCSHLWMENLIQRYIWDGESNCWSIIFYVT	960
Db	901	LLFKVLRVDTAGCLLKCSGHGCDPLTKRCITCSHLWMENLIQRYIWDGESNCWSIIFYVT	960
Qy	961	VLAFTLILVLTCGFTWLCLCCCKROKRTKIRKTKYTTILDNDMEQBERMELRPKYGIKHS	1020
Db	961	VLAFTLILVLTCGFTWLCLCCCKROKRTKIRKTKYTTILDNDMEQBERMELRPKYGIKHS	1020
Qy	1021	EHNSSLMWSESEFSDQDTIFSRKMERKGNPKVSMNGSIRNGASFYSCKDR	1072
Db	1021	EHNSSLMWSESEFSDQDTIFSRKMERKGNPKVSMNGSIRNGASFYSCKDR	1072

```

; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Arya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PiD6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1072
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-10-764-390-11

Query Match          99.9%;      Score 5576;      DB 16;      Length 1072;
Best Local Similarity 99.7%;
Matches 1069;      Conservative      0;      Mismatches      3;      Indels      0;      Gaps

```

RESULT 2  
US-10-764-390-11  
; Sequence 11, A  
; Publication No  
; GENERAL INFORM





```

RESULT 4
US-10-764-390-5
; Sequence 5, Application US/10764390
; Publication No. US2004021412A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254P1D6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1072
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-5

Query Match 99.9%; Score 5574; DB 16; Length 1072;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAPPTGVLSSLLLVTTIAGCARKQCEGRTYSNAVISPNLETRIMRVSHTEPPVVDCTAA 60
DB 1 MAPPTGVLSSLLLVTTIAGCARKQCEGRTYSNAVISPNLETRIMRVSHTEPPVVDCTAA 60
QY 61 CDLSSCDLAWFEGRCYLVSPCHKENCCKMGPIRSYLTFLVPVORPAQLLDYGDMM 120
DB 61 CDLSSCDLAWFEGRCYLVSPCHKENCCKMGPIRSYLTFLVPVORPAQLLDYGDMM 120
QY 121 LNRGSPGSGWSPEDIRKDLFLGKDWGLEEMSEYDDYRELEKDLQPSGQKPRGSA 180
DB 121 LNRGSPGSGWSPEDIRKDLFLGKDWGLEEMSEYDDYRELEKDLQPSGQKPRGSA 180
QY 181 EYTDWGLLPGSGAFNSVGDSPAVPAETQDPELHYLNESASTPAPKLPERSVLLPLPT 240
DB 181 EYTDWGLLPGSGAFNSVGDSPAVPAETQDPELHYLNESASTPAPKLPERSVLLPLPT 240
QY 241 TSSGSEVLKESKASQLOESSNSGKVLMPGSHLPPASLELSSVTVKESPVLTTPGST 300
DB 241 TSSGSEVLKESKASQLOESSNSGKVLMPGSHLPPASLELSSVTVKESPVLTTPGST 300
QY 301 EHSIPTPTSAAPSESTPSELPSPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360
DB 301 EHSIPTPTSAAPSESTPSELPSPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360
QY 361 APPVETTYEWNLSHPDYOGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAFGEFV 420
DB 361 APPVETTYEWNLSHPDYOGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAFGEFV 420
QY 421 NVTVPARRVNLPPVAVVSPQLQELTLPLTSALIDGQSQTDDTEIVSYHWEERINGPFTIE 480
DB 421 NVTVPARRVNLPPVAVVSPQLQELTLPLTSALIDGQSQTDDTEIVSYHWEERINGPFTIE 480
QY 481 KTSVDSFVLRLSNLDPGNYSFRLTVTSDGATNSTTAAALIVNNAVDPYPVANAGNHTIT 540
DB 481 KTSVDSFVLRLSNLDPGNYSFRLTVTSDGATNSTTAAALIVNNAVDPYPVANAGNHTIT 540
QY 541 LPQNSITLNGNSSDDHQIYLVSWSLGPGSEKGVVMQGVOTPYLHLSAMQEGDYYFQJK 600
DB 541 LPQNSITLNGNSSDDHQIYLVSWSLGPGSEKGVVMQGVOTPYLHLSAMQEGDYYFQJK 600

RESULT 5
US-10-764-390-8
; Sequence 8, Application US/10764390
; Publication No. US2004021412A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254P1D6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1072
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-8

Query Match 99.9%; Score 5574; DB 16; Length 1072;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAPPTGVLSSLLLVTTIAGCARKQCEGRTYSNAVISPNLETRIMRVSHTEPPVVDCTAA 60
DB 1 MAPPTGVLSSLLLVTTIAGCARKQCEGRTYSNAVISPNLETRIMRVSHTEPPVVDCTAA 60
QY 61 CDLSSCDLAWFEGRCYLVSPCHKENCCKMGPIRSYLTFLVPVORPAQLLDYGDMM 120
DB 61 CDLSSCDLAWFEGRCYLVSPCHKENCCKMGPIRSYLTFLVPVORPAQLLDYGDMM 120
QY 121 LNRGSPGSGWSPEDIRKDLFLGKDWGLEEMSEYDDYRELEKDLQPSGQKPRGSA 180
DB 121 LNRGSPGSGWSPEDIRKDLFLGKDWGLEEMSEYDDYRELEKDLQPSGQKPRGSA 180
QY 181 EYTDWGLLPGSGAFNSVGDSPAVPAETQDPELHYLNESASTPAPKLPERSVLLPLPT 240
DB 181 EYTDWGLLPGSGAFNSVGDSPAVPAETQDPELHYLNESASTPAPKLPERSVLLPLPT 240
QY 241 TSSGSEVLKESKASQLOESSNSGKVLMPGSHLPPASLELSSVTVKESPVLTTPGST 300
DB 241 TSSGSEVLKESKASQLOESSNSGKVLMPGSHLPPASLELSSVTVKESPVLTTPGST 300
QY 301 EHSIPTPTSAAPSESTPSELPSPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360
DB 301 EHSIPTPTSAAPSESTPSELPSPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360
QY 361 APPVETTYEWNLSHPDYOGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAFGEFV 420
DB 361 APPVETTYEWNLSHPDYOGEIKQGHKQTLNLSQLSVGLYVFKVTVSSNAFGEFV 420
QY 421 NVTVPARRVNLPPVAVVSPQLQELTLPLTSALIDGQSQTDDTEIVSYHWEERINGPFTIE 480
DB 421 NVTVPARRVNLPPVAVVSPQLQELTLPLTSALIDGQSQTDDTEIVSYHWEERINGPFTIE 480
QY 481 KTSVDSFVLRLSNLDPGNYSFRLTVTSDGATNSTTAAALIVNNAVDPYPVANAGNHTIT 540
DB 481 KTSVDSFVLRLSNLDPGNYSFRLTVTSDGATNSTTAAALIVNNAVDPYPVANAGNHTIT 540
QY 541 LPQNSITLNGNSSDDHQIYLVSWSLGPGSEKGVVMQGVOTPYLHLSAMQEGDYYFQJK 600
DB 541 LPQNSITLNGNSSDDHQIYLVSWSLGPGSEKGVVMQGVOTPYLHLSAMQEGDYYFQJK 600

601 VTDSSRQOSTAXVTVIQVQENNRPPVAVAGDPDKELIFPVESATLDGSSSSDDHGIVFYHW 660
601 VTDSSRQOSTAXVTVIQVQENNRPPVAVAGDPDKELIFPVESATLDGSSSSDDHGIVFYHW 660
661 EHVGPSAVEMENIDKAIATVTGLQVGYTHFRLTVKQOQGLSSSTLTVAVKENNSPPR 720
661 EHVGPSAVEMENIDKAIATVTGLQVGYTHFRLTVKQOQGLSSSTLTVAVKENNSPPR 720
721 ARAGRHVLVLPNNSITLDGSRSDTDORIVSYLWIRDOQSPAAAGDVIDGSDHSVALQLTN 780
721 ARAGRHVLVLPNNSITLDGSRSDTDORIVSYLWIRDOQSPAAAGDVIDGSDHSVALQLTN 780
781 LVEGVYTFHLRVTDQSGASDSDTATVEVQDPDKRSGLVLTQVGVGLTEQRKDTLVKQ 840
781 LVEGVYTFHLRVTDQSGASDSDTATVEVQDPDKRSGLVLTQVGVGLTEQRKDTLVKQ 840
841 LAVLLNVLDSDIKVQKIRAHSDLSSTVIIVFYQSPPPFKVLAEEVARNLHRLSKEKADF 900
841 LAVLLNVLDSDIKVQKIRAHSDLSSTVIIVFYQSPPPFKVLAEEVARNLHRLSKEKADF 900
901 LLEFVLVDTAGCLLKSGHGHCDPLTKRCICSHLWMENLIQRYIWDGESNCESWIFVYT 960
901 LLEFVLVDTAGCLLKSGHGHCDPLTKRCICSHLWMENLIQRYIWDGESNCESWIFVYT 960
961 VLAFTLIVLTGFTWLCICCCCKRQKTKIRKTKYITLDNMDEQERMELRPKYGIKHRST 1020
961 VLAFTLIVLTGFTWLCICCCCKRQKTKIRKTKYITLDNMDEQERMELRPKYGIKHRST 1020
1021 EHNSLWSESEFSDSDQDTIFSRKMERGNPKVMNGSIRNGASFYSCKDR 1072
1021 EHNSLWSESEFSDSDQDTIFSRKMERGNPKVMNGSIRNGASFYSCKDR 1072

```

```
Qy 121 LNRGSPGSGWGDSPEDIRKDLXFLGKDWGLEEMSEYXDDYRELEKDLLQPSGKQEPGRSA 180
Db 121 LNRGSPGSGWGDSPEDIRKDLXFLGKDWGLEEMSEYXDDYRELEKDLLQPSGKQEPGRSA 180
Qy 181 EYTDWGLLPGSEGAFNSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240
Db 181 EYTDWGLLPGSEGAFNSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240
Qy 241 TPSSGEVLEKASQLOEQSSNSGKVLMPSHSLPPASLELSSVTVKESPVLTVPST 300
Db 241 TPSSGEVLEKASQLOEQSSNSGKVLMPSHSLPPASLELSSVTVKESPVLTVPST 300
Qy 301 EHSIPTPTSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLITLIPDNEVELKAFVAP 360
Db 301 EHSIPTPTSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLITLIPDNEVELKAFVAP 360
Qy 361 APPVETTYNENWLI SHPTDYQGEIKQGHKQTLNLSQSLVGLYVFKVTVSSENAFGEFV 420
Db 361 APPVETTYNENWLI SHPTDYQGEIKQGHKQTLNLSQSLVGLYVFKVTVSSENAFGEFV 420
Qy 421 NVTVKPARRVNLPPVAVVSPQLBELTLP LTSALIDGSGTDDTEIVSVHWEIINGPFTIE 480
Db 421 NVTVKPARRVNLPPVAVVSPQLBELTLP LTSALIDGSGTDDTEIVSVHWEIINGPFTIE 480
Qy 481 KTSVDSVPLRLNLDPCGNYSEFRLTPTSDGATNSTTAAALIVNNAVDPVPPVANAGPNHTIT 540
Db 481 KTSVDSVPLRLNLDPCGNYSEFRLTPTSDGATNSTTAAALIVNNAVDPVPPVANAGPNHTIT 540
Qy 541 LPQNSITLNGQSSDDHQI VLYEWSLGPSEGKHVVMQGVQTPYLHL SAMQEGDYTFQLK 600
Db 541 LPQNSITLNGQSSDDHQI VLYEWSLGPSEGKHVVMQGVQTPYLHL SAMQEGDYTFQLK 600
Qy 601 VTDSSROQSTAXVTVI VQPNRRPPVAVAGDPKELI PPVESATLDGSSSSDDHGIVFYHW 660
Db 601 VTDSSROQSTAXVTVI VQPNRRPPVAVAGDPKELI PPVESATLDGSSSSDDHGIVFYHW 660
Qy 661 EHVGPASAVEMENIDKAIATVTGLQVGYTHFRLTVKQOQGLSSTSTLTVA VKKENNSPPR 720
Db 661 EHVGPASAVEMENIDKAIATVTGLQVGYTHFRLTVKQOQGLSSTSTLTVA VKKENNSPPR 720
Qy 721 ARAGGRHVLPLPNNSTLIDGSRSDTDORI VSYLWIRDOQSPAAGDVIDGSDHVALQLTN 780
Db 721 ARAGGRHVLPLPNNSTLIDGSRSDTDORI VSYLWIRDOQSPAAGDVIDGSDHVALQLTN 780
Qy 781 LVEGYTIFHLRVTD SQASD TDTATVEVQDPDKSGLVLTQVGVGQLTQQRKDTLVRQ 840
Db 781 LVEGYTIFHLRVTD SQASD TDTATVEVQDPDKSGLVLTQVGVGQLTQQRKDTLVRQ 840
Qy 841 LAVLLNVLDSI KVQKTRAHSDLS TVIVFYVQSRPPPKVLKAAEVARNLHMLRSLKEKADF 900
Db 841 LAVLLNVLDSI KVQKTRAHSDLS TVIVFYVQSRPPPKVLKAAEVARNLHMLRSLKEKADF 900
Qy 901 LLEFVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWENLIRYIWDGESNCWSIFVVT 960
Db 901 LLEFVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWENLIRYIWDGESNCWSIFVVT 960
Qy 961 VLAFTLIVLTGGFTWLCICCCCKRQKRTKIRKTKYTTI LDNNDEQERMELEPKYGI KHRST 1020
Db 961 VLAFTLIVLTGGFTWLCICCCCKRQKRTKIRKTKYTTI LDNNDEQERMELEPKYGI KHRST 1020
Qy 1021 EHNSSLVMESEFSDQDPTIFSRKMERGNPKVSNMGSIRNGASFYSCKDR 1072
Db 1021 EHNSSLVMESEFSDQDPTIFSRKMERGNPKVSNMGSIRNGASFYSCKDR 1072
```

## RESULT 6

```
US-10-764-390-9
; Sequence 9, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
```

```
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1072
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-9
```

```
Query Match 99.9%; Score 5574; DB 16; Length 1072;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 MAPPTGVLSSLLLVTTIAGCARKQCSEGRYTSNAVISPNLETTIRIMRVSHTFPVDCTAA 60
Db 1 MAPPTGVLSSLLLVTTIAGCARKQCSEGRYTSNAVISPNLETTIRIMRVSHTFPVDCTAA 60
Qy 61 CCDSLSSCDLAWPFGRCVLSVCPHKENCEPKMGPIRSYLTFLVRPQORPAQLLDYGDMM 120
Db 61 CCDSLSSCDLAWPFGRCVLSVCPHKENCEPKMGPIRSYLTFLVRPQORPAQLLDYGDMM 120
Qy 121 LNRGSPSGIWDSPEDIRKDLXFLGKDWGLEEMSEYXDDYRELEKDLLQPSGKQEPGRSA 180
Db 121 LNRGSPSGIWDSPEDIRKDLXFLGKDWGLEEMSEYXDDYRELEKDLLQPSGKQEPGRSA 180
Qy 181 EYTDWGLLPGSEGAFNSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240
Db 181 EYTDWGLLPGSEGAFNSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240
Qy 241 TPSSGEVLEKASQLOEQSSNSGKVLMPSHSLPPASLELSSVTVKESPVLTVPST 300
Db 241 TPSSGEVLEKASQLOEQSSNSGKVLMPSHSLPPASLELSSVTVKESPVLTVPST 300
Qy 301 EHSIPTPTSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLITLIPDNEVELKAFVAP 360
Db 301 EHSIPTPTSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLITLIPDNEVELKAFVAP 360
Qy 361 APPVETTYNENWLI SHPTDYQGEIKQGHKQTLNLSQSLVGLYVFKVTVSSENAFGEFV 420
Db 361 APPVETTYNENWLI SHPTDYQGEIKQGHKQTLNLSQSLVGLYVFKVTVSSENAFGEFV 420
Qy 421 NVTVKPARRVNLPPVAVVSPQLBELTLP LTSALIDGSGTDDTEIVSVHWEIINGPFTIE 480
Db 421 NVTVKPARRVNLPPVAVVSPQLBELTLP LTSALIDGSGTDDTEIVSVHWEIINGPFTIE 480
Qy 481 KTSVDSVPLRLNLDPCGNYSEFRLTPTSDGATNSTTAAALIVNNAVDPVPPVANAGPNHTIT 540
Db 481 KTSVDSVPLRLNLDPCGNYSEFRLTPTSDGATNSTTAAALIVNNAVDPVPPVANAGPNHTIT 540
Qy 541 LPQNSITLNGQSSDDHQI VLYEWSLGPSEGKHVVMQGVQTPYLHL SAMQEGDYTFQLK 600
Db 541 LPQNSITLNGQSSDDHQI VLYEWSLGPSEGKHVVMQGVQTPYLHL SAMQEGDYTFQLK 600
Qy 601 VTDSSROQSTAXVTVI VQPNRRPPVAVAGDPKELI PPVESATLDGSSSSDDHGIVFYHW 660
Db 601 VTDSSROQSTAXVTVI VQPNRRPPVAVAGDPKELI PPVESATLDGSSSSDDHGIVFYHW 660
Qy 661 EHVGPASAVEMENIDKAIATVTGLQVGYTHFRLTVKQOQGLSSTSTLTVA VKKENNSPPR 720
Db 661 EHVGPASAVEMENIDKAIATVTGLQVGYTHFRLTVKQOQGLSSTSTLTVA VKKENNSPPR 720
```

```

QY 721 ARAGGRHVLVLPNNSITLDGSRSTDDQRIYSYLWIRDGQSPAAGDVIDGSDHSHVALQLTN 780
Db 721 ARAGGRHVLVLPNNSITLDGSRSTDDQRIYSYLWIRDGQSPAAGDVIDGSDHSHVALQLTN 780
QY 781 LVEGYVTHLRLVTDGSGASDITATVEVQDPKRSGLVELTLQVGVGQLTQQRKDTLVRQ 840
Db 781 LVEGYVTHLRLVTDGSGASDITATVEVQDPKRSGLVELTLQVGVGQLTQQRKDTLVRQ 840
QY 841 LAVLLNVLDSDIKVKIRAHSDLSITVIVYVQSRPPFKVLAEEVARNLHMRLSKEKADF 900
Db 841 LAVLLNVLDSDIKVKIRAHSDLSITVIVYVQSRPPFKVLAEEVARNLHMRLSKEKADF 900
QY 901 LFFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCESIFVYT 960
Db 901 LFFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCESIFVYT 960
QY 961 VLAFTLIVLTGGFTWLCICCCCKRQRTKIRKTKYTILDNMDEQRMELRPKYGIKHRST 1020
Db 961 VLAFTLIVLTGGFTWLCICCCCKRQRTKIRKTKYTILDNMDEQRMELRPKYGIKHRST 1020
QY 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKOR 1072
Db 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKOR 1072

RESULT 7
US-10-764-390-12
; Sequence 12, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1D6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1072
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-764-390-12

Query Match 99.9%; Score 5574; DB 16; Length 1072;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAPPTGVLSSLLLVITAGCARQCSEGRTYNAVISPNELETTRIMRVSHHTPPVVDCTAA 60
Db 1 MAPPTGVLSSLLLVITAGCARQCSEGRTYNAVISPNELETTRIMRVSHHTPPVVDCTAA 60
QY 61 CDDLSLSCDLAWFEGRCYLVSCHPKNCEPKWGPPIRSYLTFLVRVQRPQALLDYGDMM 120
Db 61 CDDLSLSCDLAWFEGRCYLVSCHPKNCEPKWGPPIRSYLTFLVRVQRPQALLDYGDMM 120
QY 121 LNRGSPGIWGSPEIRKDLPLGKDWGLEEMSEYXDDYRELEKDLQPSGKQPRGSA 180
Db 121 LNRGSPGIWGSPEIRKDLPLGKDWGLEEMSEYADYRELEKDLQPSGKQPRGSA 180
QY 181 EYTDWGLLPSEGAFNSSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240
Db 181 EYTDWGLLPSEGAFNSSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240

```

```

Db 181 EYTDWGLLPSEGAFNSSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPT 240
QY 241 TPSSGEVLEKEKASQLOEQSSNSGKEVLMPSHSLPPASLELSVTVTEKSPVLTVTPGST 300
Db 241 TPSSGEVLEKEKASQLOEQSSNSGKEVLMPSHSLPPASLELSVTVTEKSPVLTVTPGST 300
QY 301 EHSIPTPTGSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360
Db 301 EHSIPTPTGSAAPSESTPSELPISTPTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAP 360
QY 361 APPVETTYNTEWNLISHTDYQGEIKOQHKOTLNLSQLSVGLYVFKVTVSSENAFGEQFV 420
Db 361 APPVETTYNTEWNLISHTDYQGEIKOQHKOTLNLSQLSVGLYVFKVTVSSENAFGEQFV 420
QY 421 NVTVKPARRVNLPPVAVVSPQLQELTLPLTSALIDGSGSTDDTEIVSYHWEIINGPFIEE 480
Db 421 NVTVKPARRVNLPPVAVVSPQLQELTLPLTSALIDGSGSTDDTEIVSYHWEIINGPFIEE 480
QY 481 KTSVDSVPLRLSNLDPNGYSFRLTVTSDGATNSTTAAALIVNNAVDPYPPVANAGPNHTIT 540
Db 481 KTSVDSVPLRLSNLDPNGYSFRLTVTSDGATNSTTAAALIVNNAVDPYPPVANAGPNHTIT 540
QY 541 LPQNSITLNGQSSDDHQIIVLYEWSLPGSGRKHVVMQGVQTPYLHLSAMQEGDITFOLK 600
Db 541 LPQNSITLNGQSSDDHQIIVLYEWSLPGSGRKHVVMQGVQTPYLHLSAMQEGDITFOLK 600
QY 601 VTDSRQOSTAXVTVIIVQPENNRPPVAVAGDPKELIPPVESATLDGSSSSDDHGIVFYHW 660
Db 601 VTDSRQOSTAXVTVIIVQPENNRPPVAVAGDPKELIPPVESATLDGSSSSDDHGIVFYHW 660
QY 661 EHVGRPSAVEMENTDKAIAITVGLQVGYTHFRLTVKQOQGLSSTSTLTVAKKENNSPPR 720
Db 661 EHVGRPSAVEMENTDKAIAITVGLQVGYTHFRLTVKQOQGLSSTSTLTVAKKENNSPPR 720
QY 721 ARAGGRHVLVLPNNSITLDGSRSTDDQRIYSYLWIRDGQSPAAGDVIDGSDHSHVALQLTN 780
Db 721 ARAGGRHVLVLPNNSITLDGSRSTDDQRIYSYLWIRDGQSPAAGDVIDGSDHSHVALQLTN 780
QY 781 LVEGYVTHLRLVTDGSGASDITATVEVQDPKRSGLVELTLQVGVGQLTQQRKDTLVRQ 840
Db 781 LVEGYVTHLRLVTDGSGASDITATVEVQDPKRSGLVELTLQVGVGQLTQQRKDTLVRQ 840
QY 841 LAVLLNVLDSDIKVKIRAHSDLSITVIVYVQSRPPFKVLAEEVARNLHMRLSKEKADF 900
Db 841 LAVLLNVLDSDIKVKIRAHSDLSITVIVYVQSRPPFKVLAEEVARNLHMRLSKEKADF 900
QY 901 LFFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCESIFVYT 960
Db 901 LFFKVLRLVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIWDGESNCESIFVYT 960
QY 961 VLAFTLIVLTGGFTWLCICCCCKRQRTKIRKTKYTILDNMDEQRMELRPKYGIKHRST 1020
Db 961 VLAFTLIVLTGGFTWLCICCCCKRQRTKIRKTKYTILDNMDEQRMELRPKYGIKHRST 1020
QY 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKOR 1072
Db 1021 EHNSLMVSESEFSDQDTIFSRKMERGNPKVSMNGSIRNGASFSYCSKOR 1072

RESULT 8
US-10-764-390-259
; Sequence 259, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins

```

; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of  
 ; TITLE OF INVENTION: Cancer  
 ; FILE REFERENCE: 51158-20081.00  
 ; CURRENT APPLICATION NUMBER: US/10/764,390  
 ; PRIOR FILING DATE: 2004-01-23  
 ; PRIOR APPLICATION NUMBER: US60/442,526  
 ; PRIOR FILING DATE: 2003-01-24  
 ; NUMBER OF SEQ ID NOS: 277  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 259  
 ; LENGTH: 1072  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-764-390-259

Query Match 99.9%; Score 5574; DB 16; Length 1072;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MAPTGVLSLLLLVTTIAGCARKQCSGRTYTSNAVISPNLETTIRMRVSHTFPPVVDCTAA	60
Db	1	MAPTGVLSLLLLVTTIAGCARKQCSGRTYTSNAVISPNLETTIRMRVSHTFPPVVDCTAA	60
Qy	61	CCDLSSCDLAWFEGRCYLVSPCHKENCEPKMGPIRSYLTFLVRPQVPAQLLDYDMM	120
Db	61	CCDLSSCDLAWFEGRCYLVSPCHKENCEPKMGPIRSYLTFLVRPQVPAQLLDYDMM	120
Qy	121	LNRRSGPSGIDSPEDIRKDLFLGKOWGLEEMSEYDXYRELEKDLLQPSGKQPRGSA	180
Db	121	LNRRSGPSGIDSPEDIRKDLFLGKOWGLEEMSEYDXYRELEKDLLQPSGKQPRGSA	180
Qy	181	EYTDWGLLPQSGEAFNSVGSVPAPVAPQDQDELHYLINESASTPAPKLPERSVLLPLPT	240
Db	181	EYTDWGLLPQSGEAFNSVGSVPAPVAPQDQDELHYLINESASTPAPKLPERSVLLPLPT	240
Qy	241	TPSSEVLEKESAKOLQOSNSGKVLMPSHSLPPASLELSVTVVEKSPVLTTPGST	300
Db	241	TPSSEVLEKESAKOLQOSNSGKVLMPSHSLPPASLELSVTVVEKSPVLTTPGST	300
Qy	301	EHSIPTPTSAAPSESTPSELPISPTTAPRTVKELTVSAGDNLITLDPNEVELKAFVAP	360
Db	301	EHSIPTPTSAAPSESTPSELPISPTTAPRTVKELTVSAGDNLITLDPNEVELKAFVAP	360
Qy	361	APPVETTYNENWNLISHPTDYQGEIKQGHKOTLNLSQLSVGLYFVKTVSSENAFGGFV	420
Db	361	APPVETTYNENWNLISHPTDYQGEIKQGHKOTLNLSQLSVGLYFVKTVSSENAFGGFV	420
Qy	421	NVTVKPARRVNLPPVAVVSPQLBELTLPALSALDGSQSTDDTEIVSYHWEINGPTEE	480
Db	421	NVTVKPARRVNLPPVAVVSPQLBELTLPALSALDGSQSTDDTEIVSYHWEINGPTEE	480
Qy	481	KTSVDSPLRLSNLDPGNYSFRLTVDSDGATNSTTAALIVNNAVDYPPVANAGPNHTIT	540
Db	481	KTSVDSPLRLSNLDPGNYSFRLTVDSDGATNSTTAALIVNNAVDYPPVANAGPNHTIT	540
Qy	541	LPQNSITLNGQSDDDHQIVLYENSLGPGSGKHVMQGVQTPYLHLSAMQEGDYTFQLK	600
Db	541	LPQNSITLNGQSDDDHQIVLYENSLGPGSGKHVMQGVQTPYLHLSAMQEGDYTFQLK	600
Qy	601	VTDSSRQSTAXVTIVOPENNRPPVAVAGDPKELIPVESATLDGSSSDDHGIVFVHW	660
Db	601	VTDSSRQSTAXVTIVOPENNRPPVAVAGDPKELIPVESATLDGSSSDDHGIVFVHW	660
Qy	661	EHVGPASAVENIDKALATVTLQVGYTHFLRTVKDQOGLSSTTLTVAVKKNNSPPR	720
Db	661	EHVGPASAVENIDKALATVTLQVGYTHFLRTVKDQOGLSSTTLTVAVKKNNSPPR	720
Qy	721	ARAGRHVLPLNNSITLDGSRDQDQIVSYLWIRDCQSPAAGDVIDGSDHVALQLTN	780
Db	721	ARAGRHVLPLNNSITLDGSRDQDQIVSYLWIRDCQSPAAGDVIDGSDHVALQLTN	780
Qy	781	LVEGYTHFLRVTDQSGASDQDTATVEVQDPPKSGLVGLVELTLQVGVQLTQKDTLVQ	840

Db	781	LVEGYTHFLRVTDQSGASDQDTATVEVQDPPKSGLVGLVELTLQVGVQLTQKDTLVQ	840
Qy	841	LAVLLNVLDSDIKVQKIRAHSDLSSTVIVFYVQSRPPFKVLKAAEVARNLHMLSKKADP	900
Db	841	LAVLLNVLDSDIKVQKIRAHSDLSSTVIVFYVQSRPPFKVLKAAEVARNLHMLSKKADP	900
Qy	901	LLPKVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIMDGESNCESIFVVT	960
Db	901	LLPKVLRVDTAGCLLKCSGHGCHDPLTKRCICSHLWMENLIQRYIMDGESNCESIFVVT	960
Qy	961	VLAFTLIVLTGGFTWLCICCCCKRQRTKIRKTKYTILDNMDEQERMELRKYGIKHST	1020
Db	961	VLAFTLIVLTGGFTWLCICCCCKRQRTKIRKTKYTILDNMDEQERMELRKYGIKHST	1020
Qy	1021	EHNSLMVSESEFSDQDTIFSRKMERGNPKVMNGSIRNGASFVCSKDR	1072
Db	1021	EHNSLMVSESEFSDQDTIFSRKMERGNPKVMNGSIRNGASFVCSKDR	1072

RESULT 9  
 US-10-764-390-273  
 ; Sequence 273, Application US/10764390  
 ; Publication No. US20040214212A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Agensys, Inc.  
 ; APPLICANT: Raitano, Arthur B.  
 ; APPLICANT: Jakobovits, Aya  
 ; APPLICANT: Challita-Bid, Pia M.  
 ; APPLICANT: Ge, Wangmao  
 ; APPLICANT: Paris, Mary  
 ; APPLICANT: Steven B. Kanner  
 ; APPLICANT: Juan J. Perez-Villar  
 ; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
 ; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of  
 ; TITLE OF INVENTION: Cancer  
 ; FILE REFERENCE: 51158-20081.00  
 ; CURRENT APPLICATION NUMBER: US/10/764,390  
 ; PRIOR FILING DATE: 2004-01-23  
 ; PRIOR APPLICATION NUMBER: US60/442,526  
 ; PRIOR FILING DATE: 2003-01-24  
 ; NUMBER OF SEQ ID NOS: 277  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 273  
 ; LENGTH: 1053  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-764-390-273

Query Match 98.3%; Score 5485; DB 16; Length 1053;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	20	CARKQCSGRTYTSNAVISPNLETTIRMRVSHTFPPVVDCTAACCDLSSCDLAWFEGRCYL	79
Db	1	CARKQCSGRTYTSNAVISPNLETTIRMRVSHTFPPVVDCTAACCDLSSCDLAWFEGRCYL	60
Qy	80	VSCPKHNCPCPKMGPIRSYLTFLVRPQVPAQLLDYDGMMLNRGSPSGIWDSPEDIRK	139
Db	61	VSCPKHNCPCPKMGPIRSYLTFLVRPQVPAQLLDYDGMMLNRGSPSGIWDSPEDIRK	120
Qy	140	DLXPLGKOWGLEEMSEYDXYRELEKOLLQPSGKQPRGSAEYTDWGLLPQSGEAFNSV	199
Db	121	DLXPLGKOWGLEEMSEYDXYRELEKOLLQPSGKQPRGSAEYTDWGLLPQSGEAFNSV	180
Qy	200	GDSPVAPAEQDDELHYLINESASTPAPKLPERSVLLPLPTTPSSGVELEKEKASQOEQ	259
Db	181	GDSPVAPAEQDDELHYLINESASTPAPKLPERSVLLPLPTTPSSGVELEKEKASQOEQ	240
Qy	260	SSNSGKVLMPSHSLPPASLELSVTVVEKSPVLTTPGSTEHSIPTPTSAAPSESTPS	319
Db	241	SSNSGKVLMPSHSLPPASLELSVTVVEKSPVLTTPGSTEHSIPTPTSAAPSESTPS	300
Qy	320	ELPISPTTAPRTVKELTVSAGDNLITLDPNEVELKAFVAPVETTYNENWNLISHPT	379

```
Db 301 ELPISTPTAPRTVKELTVSAGDNLIITLPDNEVELKAFVADAPPVETTYNENLISHPT 360
Qy 380 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTYSSENAFGEFVNVTVPARRVNLPPVAVVS 439
Db 361 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTYSSENAFGEFVNVTVPARRVNLPPVAVVS 420
Qy 440 POLQELTLPLTSALIDGSGTDDTEIVSYHWEENGPFIIEKTSVDSVPLRLSNLDPGNY 499
Db 421 POLQELTLPLTSALIDGSGTDDTEIVSYHWEENGPFIIEKTSVDSVPLRLSNLDPGNY 480
Qy 500 SFRLTVTDSGATNSTTAALIVNNAVDPYPVANAGPNHTITLPQNSITLNGQSSDDHQI 559
Db 481 SFRLTVTDSGATNSTTAALIVNNAVDPYPVANAGPNHTITLPQNSITLNGQSSDDHQI 540
Qy 560 VLYEWSLGPSEGKHVVMQGVQTPYLHLISAMQEGDYTFQLKVTDSRQOSTAXVTIVQ 619
Db 541 VLYEWSLGPSEGKHVVMQGVQTPYLHLISAMQEGDYTFQLKVTDSRQOSTAXVTIVQ 600
Qy 620 ENNRPPVAVAGDPKELIFPVESATLDGSSSDHGI VFYHWEHVRGSAVEMENIDKAIA 679
Db 601 ENNRPPVAVAGDPKELIFPVESATLDGSSSDHGI VFYHWEHVRGSAVEMENIDKAIA 660
Qy 680 TVTGLQVGYTHFRLTVKDOQGLSSTLTVAVKKENNSPPRARAGRHVLPNNISITLD 739
Db 661 TVTGLQVGYTHFRLTVKDOQGLSSTLTVAVKKENNSPPRARAGRHVLPNNISITLD 720
Qy 740 GSRSTDDQRIIVSYLWIRDOQSPAAGDVTDGSDHVALQTLNVEGYTFHLRVTDSDQAS 799
Db 721 GSRSTDDQRIIVSYLWIRDOQSPAAGDVTDGSDHVALQTLNVEGYTFHLRVTDSDQAS 780
Qy 800 DTDATVEVQDPKRSGLVELTLQVGVQLTEQRKDTLVRQLAVLNVLDSIKVQKIRA 859
Db 781 DTDATVEVQDPKRSGLVELTLQVGVQLTEQRKDTLVRQLAVLNVLDSIKVQKIRA 840
Qy 860 HSDLSSTVIVFYVQSRPPPKVLKAAEVARNLHMLRSLKEKADFLFKVLRVDTAGCLLKCSG 919
Db 841 HSDLSSTVIVFYVQSRPPPKVLKAAEVARNLHMLRSLKEKADFLFKVLRVDTAGCLLKCSG 900
Qy 920 HGHCDBLTKRCICSHLWENLQRYIWDGESNCEWSIFVYTVLAFTLIVLTGFTWLCIC 979
Db 901 HGHCDBLTKRCICSHLWENLQRYIWDGESNCEWSIFVYTVLAFTLIVLTGFTWLCIC 960
Qy 980 CCKRQKTKIRKTKYITLDNDWDEQRMELRPKYGIKHRSTENHSLMVSESEFSDQDT 1039
Db 961 CCKRQKTKIRKTKYITLDNDWDEQRMELRPKYGIKHRSTENHSLMVSESEFSDQDT 1020
Qy 1040 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1072
Db 1021 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1053
```

```
RESULT 10
US-10-764-390-274
; Sequence 274, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254P1d68 Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
```

```
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 274
; LENGTH: 1053
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-274

Query Match          98.3%; Score 5485; DB 16; Length 1053;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 20 CARQCSGRTYSNAVISPNLETTRIMRVSHPTPVDCCTAACCDLSSCDLAWFEGRCYL 79
Db 1 CARQCSGRTYSNAVISPNLETTRIMRVSHPTPVDCCTAACCDLSSCDLAWFEGRCYL 60

Qy 80 VSCPHKENCERPKMGPIRSYLTFLVRPVRPAQLDLYCDMMLNKSGPSG1WGDSPEDIRK 139
Db 61 VSCPHKENCERPKMGPIRSYLTFLVRPVRPAQLDLYCDMMLNKSGPSG1WGDSPEDIRK 120

Qy 140 DLXFLGKDWGLEEMSEYKDDYRELEKOLLQPSGKQEPGSAEYTDWGLLPGESEAFNSV 199
Db 121 DLPFLGKDWGLEEMSEYKDDYRELEKOLLQPSGKQEPGSAEYTDWGLLPGESEAFNSV 180

Qy 200 GDSPAVPAETQODPELHYLINESASTPAPKLPERSVLLPLPTTPSSGEVLEKEKASQLEQ 259
Db 181 GDSPAVPAETQODPELHYLINESASTPAPKLPERSVLLPLPTTPSSGEVLEKEKASQLEQ 240

Qy 260 SSNSGSEKVLMPHSLPPASLELSSVTVEKSPVLTVTPGSTEHSITPTPTSAASESTPS 319
Db 241 SSNSGSEKVLMPHSLPPASLELSSVTVEKSPVLTVTPGSTEHSITPTPTSAASESTPS 300

Qy 320 ELPISPTTAPTPTVKELTVSAGDNLIITLPDNEVELKAFVADAPPVETTYNENLISHPT 379
Db 301 ELPISPTTAPTPTVKELTVSAGDNLIITLPDNEVELKAFVADAPPVETTYNENLISHPT 360

Qy 380 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTYSSENAFGEFVNVTVPARRVNLPPVAVVS 439
Db 361 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTYSSENAFGEFVNVTVPARRVNLPPVAVVS 420

Qy 440 POLQELTLPLTSALIDGSGTDDTEIVSYHWEENGPFIIEKTSVDSVPLRLSNLDPGNY 499
Db 421 POLQELTLPLTSALIDGSGTDDTEIVSYHWEENGPFIIEKTSVDSVPLRLSNLDPGNY 480

Qy 500 SFRLTVTDSGATNSTTAALIVNNAVDPYPVANAGPNHTITLPQNSITLNGQSSDDHQI 559
Db 481 SFRLTVTDSGATNSTTAALIVNNAVDPYPVANAGPNHTITLPQNSITLNGQSSDDHQI 540

Qy 560 VLYEWSLGPSEGKHVVMQGVQTPYLHLISAMQEGDYTFQLKVTDSRQOSTAXVTIVQ 619
Db 541 VLYEWSLGPSEGKHVVMQGVQTPYLHLISAMQEGDYTFQLKVTDSRQOSTAXVTIVQ 600

Qy 620 ENNRPPVAVAGDPKELIFPVESATLDGSSSDHGI VFYHWEHVRGSAVEMENIDKAIA 679
Db 601 ENNRPPVAVAGDPKELIFPVESATLDGSSSDHGI VFYHWEHVRGSAVEMENIDKAIA 660

Qy 680 TVTGLQVGYTHFRLTVKDOQGLSSTLTVAVKKENNSPPRARAGRHVLPNNISITLD 739
Db 661 TVTGLQVGYTHFRLTVKDOQGLSSTLTVAVKKENNSPPRARAGRHVLPNNISITLD 720

Qy 740 GSRSTDDQRIIVSYLWIRDOQSPAAGDVTDGSDHVALQTLNVEGYTFHLRVTDSDQAS 799
Db 721 GSRSTDDQRIIVSYLWIRDOQSPAAGDVTDGSDHVALQTLNVEGYTFHLRVTDSDQAS 780

Qy 800 DTDATVEVQDPKRSGLVELTLQVGVQLTEQRKDTLVRQLAVLNVLDSIKVQKIRA 859
Db 781 DTDATVEVQDPKRSGLVELTLQVGVQLTEQRKDTLVRQLAVLNVLDSIKVQKIRA 840

Qy 860 HSDLSSTVIVFYVQSRPPPKVLKAAEVARNLHMLRSLKEKADFLFKVLRVDTAGCLLKCSG 919
Db 841 HSDLSSTVIVFYVQSRPPPKVLKAAEVARNLHMLRSLKEKADFLFKVLRVDTAGCLLKCSG 900

Qy 920 HGHCDBLTKRCICSHLWENLQRYIWDGESNCEWSIFVYTVLAFTLIVLTGFTWLCIC 979
Db 901 HGHCDBLTKRCICSHLWENLQRYIWDGESNCEWSIFVYTVLAFTLIVLTGFTWLCIC 960

Qy 980 CCKRQKTKIRKTKYITLDNDWDEQRMELRPKYGIKHRSTENHSLMVSESEFSDQDT 1039
Db 961 CCKRQKTKIRKTKYITLDNDWDEQRMELRPKYGIKHRSTENHSLMVSESEFSDQDT 1020

Qy 1040 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1072
Db 1021 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1053
```



```

Db 901 HGHCDPLTKRCICSHLWMENLIQRYIWDGSSNCWSIFYYTVLAFTLIVTGGTWCIC 960
Qy 980 CCKRQKTKIRKTKYTIIDNMDEQERMELEPKYGIKHSRSTEHNSLWSESEFSDSDT 1039
Db 961 CCKRQKTKIRKTKYTIIDNMDEQERMELEPKYGIKHSRSTEHNSLWSESEFSDSDT 1020
Qy 1040 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1072
Db 1021 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1053

RESULT 11
US-10-764-390-7
; Sequence 7, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Faris, Mary
; APPLICANT: Ge, Wangmao
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; FILE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1063
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-7

Query Match 98.38; Score 5485; DB 16; Length 1063;
Best Local Similarity 99.74; Pred. No. 0;
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 20 CARQKQCEGRTYSNAVISPNLETRIMRVSHTFPVVDCTAACCDLSSCDLAWFEGRCYL 79
Db 11 CARQKQCEGRTYSNAVISPNLETRIMRVSHTFPVVDCTAACCDLSSCDLAWFEGRCYL 70
Qy 80 VSCPHKENCBPQKMGPIRSYLTFLVRPQPAQLLDYGDMLNRGSPSGIWGDSPEDIRK 139
Db 71 VSCPHKENCBPQKMGPIRSYLTFLVRPQPAQLLDYGDMLNRGSPSGIWGDSPEDIRK 130
Qy 140 DLXFLGKDWGLEMSEYXDYRELEKDLQPSGKQEPGRGAEYTDWGLLPGSEGAFFNSV 199
Db 131 DLXFLGKDWGLEMSEYXDYRELEKDLQPSGKQEPGRGAEYTDWGLLPGSEGAFFNSV 190
Qy 200 GDSFAVPAETQDDPELHYLNESASTPAKLPERSVLLPLTPPSSGVELEKEKASQLEQ 259
Db 191 GDSFAVPAETQDDPELHYLNESASTPAKLPERSVLLPLTPPSSGVELEKEKASQLEQ 250
Qy 260 SSNSGKEVLMPSHSLPPASLELSVTVKESPVLTTPGSTEHSIPTPTSAAPSESTPS 319
Db 251 SSNSGKEVLMPSHSLPPASLELSVTVKESPVLTTPGSTEHSIPTPTSAAPSESTPS 310
Qy 320 ELPISTPTARTVKELTVSAGDNLIIITLPDNEVELKAFAPAPPVETTYNEMNLISHT 379
Db 311 ELPISTPTARTVKELTVSAGDNLIIITLPDNEVELKAFAPAPPVETTYNEMNLISHT 370
Qy 380 DYQGEIKQGHKOTLNLSQLSGLVFKVTVSSNAFGEFVNVTVPKARRVNLPPVAVS 439
Db 371 DYQGEIKQGHKOTLNLSQLSGLVFKVTVSSNAFGEFVNVTVPKARRVNLPPVAVS 430

```

```

Qy 440 POLQELTLPLTSALIDGSSQSTDDTEIVSYHWEINGPFIIEKTSVDSPLVLSNLDPGNY 499
Db 431 POLQELTLPLTSALIDGSSQSTDDTEIVSYHWEINGPFIIEKTSVDSPLVLSNLDPGNY 490
Qy 500 SFRLTVTSDGATNSTTAALI VNNADVPPVANAGPNHTITLPONSTITLNGNQSSDDHQI 559
Db 491 SFRLTVTSDGATNSTTAALI VNNADVPPVANAGPNHTITLPONSTITLNGNQSSDDHQI 550
Qy 560 VLYEWSLGPSEKGVHVMQVQTPYLHLSAMQEGDYTFQLKVTDSSRQQSTAXVTVIVQP 619
Db 551 VLYEWSLGPSEKGVHVMQVQTPYLHLSAMQEGDYTFQLKVTDSSRQQSTAXVTVIVQP 610
Qy 620 ENNRPPVAVAGDPKELIFPVESATLDGSSSDDDHGI VFYHWEHVGRGSAVEMENIDKAIA 679
Db 611 ENNRPPVAVAGDPKELIFPVESATLDGSSSDDDHGI VFYHWEHVGRGSAVEMENIDKAIA 670
Qy 680 TVTGLQVGYTHFRLTVKDOQGLSSTLTVAVKKENNSPPRARAGGRHVLVLPNNSITLD 739
Db 671 TVTGLQVGYTHFRLTVKDOQGLSSTLTVAVKKENNSPPRARAGGRHVLVLPNNSITLD 730
Qy 740 GSRSTDQRIVSYLWIRDOGSPAAGDVIDGSDHSVALQNLNVEGVYTFHLRVTDSSQAS 799
Db 731 GSRSTDQRIVSYLWIRDOGSPAAGDVIDGSDHSVALQNLNVEGVYTFHLRVTDSSQAS 790
Qy 800 DTDATVVEQDPDRKSLVELTLQVGVGQLTQEKOTLVRQLAVLLNVLSDIKVQKIRA 859
Db 791 DTDATVVEQDPDRKSLVELTLQVGVGQLTQEKOTLVRQLAVLLNVLSDIKVQKIRA 850
Qy 860 HSDLSTVIVFYVQSRPPFKVLKAAEVARNLHMLSKADFLFKVLRVDTAGCLLKCSG 919
Db 851 HSDLSTVIVFYVQSRPPFKVLKAAEVARNLHMLSKADFLFKVLRVDTAGCLLKCSG 910
Qy 920 HGHCDPLTKRCICSHLWMENLIQRYIWDGSSNCWSIFYYTVLAFTLIVTGGTWCIC 979
Db 911 HGHCDPLTKRCICSHLWMENLIQRYIWDGSSNCWSIFYYTVLAFTLIVTGGTWCIC 970
Qy 980 CCKRQKTKIRKTKYTIIDNMDEQERMELEPKYGIKHSRSTEHNSLWSESEFSDSDT 1039
Db 971 CCKRQKTKIRKTKYTIIDNMDEQERMELEPKYGIKHSRSTEHNSLWSESEFSDSDT 1030
Qy 1040 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1072
Db 1031 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1063

RESULT 12
US-10-764-390-10
; Sequence 10, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; FILE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1063
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-10

```

```
Query Match      98.3%; Score 5485; DB 16; Length 1063;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 20 CARQCSGRTYSNAVISPNLETTIRMRVSHFTFPVVDCTAACCDLSSCDLAWFEGRCYL 79
Db 11 CARQCSGRTYSNAVISPNLETTIRMRVSHFTFPVVDCTAACCDLSSCDLAWFEGRCYL 70
Qy 80 VSCPHKENCEPKMGPIRSYLTFLVRPVQRPQAQLLDYGDMLNRGSPSGIWDSPEDIRK 139
Db 71 VSCPHKENCEPKMGPIRSYLTFLVRPVQRPQAQLLDYGDMLNRGSPSGIWDSPEDIRK 130
Qy 140 DLXFLGKDWGLEEMSEYKDDYRELEKOLLQPSGKQEPGSAEYTDWGLLPSEGFNFSV 199
Db 131 DLFLGKDWGLEEMSEYKDDYRELEKOLLQPSGKQEPGSAEYTDWGLLPSEGFNFSV 190
Qy 200 GDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPTTPSGTEHSIPTPTSAAPSESTPS 259
Db 191 GDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPTTPSGTEHSIPTPTSAAPSESTPS 250
Qy 260 SSNSGKEVLMPSHSLPPASLELSVTVKESPVLTFTPGSTEHSIPTPTSAAPSESTPS 319
Db 251 SSNSGKEVLMPSHSLPPASLELSVTVKESPVLTFTPGSTEHSIPTPTSAAPSESTPS 310
Qy 320 ELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPAPPVETTYNVEWNLISHPT 379
Db 311 ELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPAPPVETTYNVEWNLISHPT 370
Qy 380 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSENAFGEFNVTVKPARRVNLPVAVVS 439
Db 371 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSENAFGEFNVTVKPARRVNLPVAVVS 430
Qy 440 POLQELTLPLTSALIDGSGQSTDDETEIVSYHWEENGINEPIEEKTSVDSVPLRLSNLDPGNY 499
Db 431 POLQELTLPLTSALIDGSGQSTDDETEIVSYHWEENGINEPIEEKTSVDSVPLRLSNLDPGNY 490
Qy 500 SFRLTVTDSGATNSTTAALI VNNADVPPVANAGPNHTITLPQNSITLNGQSSDDHQI 559
Db 491 SFRLTVTDSGATNSTTAALI VNNADVPPVANAGPNHTITLPQNSITLNGQSSDDHQI 550
Qy 560 VLYEWSLGFSGSGKHVMQGVQTPYLHLISAMQEGDYTFQLKVTDSRQSQSTAXVTVIQVP 619
Db 551 VLYEWSLGFSGSGKHVMQGVQTPYLHLISAMQEGDYTFQLKVTDSRQSQSTAXVTVIQVP 610
Qy 620 ENRPPVAVAGDPKELIIPVESATLDGSSSDDHGIVFYHWEHVRGSAVEMENIDKAIA 679
Db 611 ENRPPVAVAGDPKELIIPVESATLDGSSSDDHGIVFYHWEHVRGSAVEMENIDKAIA 670
Qy 680 TVTGLQVGYHPRLTVKQOQGLSSTLTVA VKENNSPPRARAGRHVLVLPNNSITLD 739
Db 671 TVTGLQVGYHPRLTVKQOQGLSSTLTVA VKENNSPPRARAGRHVLVLPNNSITLD 730
Qy 740 GSRSTDDQRIVSYLIRDCQSPAAGDVIDGSDHVALQTLNLVEGYTTFHLRVTDSQAS 799
Db 731 GSRSTDDQRIVSYLIRDCQSPAAGDVIDGSDHVALQTLNLVEGYTTFHLRVTDSQAS 790
Qy 800 DTDATAVEQDPKRGSLVELTLQVGVQLTQQRKDTLVRQLAVLNVLDSIDKVKQIRA 859
Db 791 DTDATAVEQDPKRGSLVELTLQVGVQLTQQRKDTLVRQLAVLNVLDSIDKVKQIRA 850
Qy 860 HSDLSVTIVFYVQSRPPPKVLAEEVARNLHRLSKEKADFLFKVLRVDTAGCLLKCSG 919
Db 851 HSDLSVTIVFYVQSRPPPKVLAEEVARNLHRLSKEKADFLFKVLRVDTAGCLLKCSG 910
Qy 920 HGHCDPLTKRCICSHLWMENLIQRYIWDGESNCWSIFVYTVLAFTLIVLTGFTWLCIC 979
Db 911 HGHCDPLTKRCICSHLWMENLIQRYIWDGESNCWSIFVYTVLAFTLIVLTGFTWLCIC 970
Qy 980 CCKRQKRTKIRKTKYITLDNMDQERBELRPKYGIKHRSTHNSSLMWSESEFSDQDT 1039
Db 971 CCKRQKRTKIRKTKYITLDNMDQERBELRPKYGIKHRSTHNSSLMWSESEFSDQDT 1030
```

```
Qy 1040 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1072
Db 1031 IFSREKMERGNPKVSMNGSIRNGASFYSYCSKDR 1063

RESULT 13
US-10-764-390-272
; Sequence 272, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge. Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10764.390
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442.526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272
; LENGTH: 1063
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-272

Query Match      98.3%; Score 5485; DB 16; Length 1063;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1050; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 20 CARQCSGRTYSNAVISPNLETTIRMRVSHFTFPVVDCTAACCDLSSCDLAWFEGRCYL 79
Db 11 CARQCSGRTYSNAVISPNLETTIRMRVSHFTFPVVDCTAACCDLSSCDLAWFEGRCYL 70
Qy 80 VSCPHKENCEPKMGPIRSYLTFLVRPVQRPQAQLLDYGDMLNRGSPSGIWDSPEDIRK 139
Db 71 VSCPHKENCEPKMGPIRSYLTFLVRPVQRPQAQLLDYGDMLNRGSPSGIWDSPEDIRK 130
Qy 140 DLXFLGKDWGLEEMSEYKDDYRELEKOLLQPSGKQEPGSAEYTDWGLLPSEGFNFSV 199
Db 131 DLFLGKDWGLEEMSEYKDDYRELEKOLLQPSGKQEPGSAEYTDWGLLPSEGFNFSV 190
Qy 200 GDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPTTPSGTEHSIPTPTSAAPSESTPS 259
Db 191 GDSPAVPAETQDDPELHYLNESASTPAPKLPERSVLLPLPTTPSGTEHSIPTPTSAAPSESTPS 250
Qy 260 SSNSGKEVLMPSHSLPPASLELSVTVKESPVLTFTPGSTEHSIPTPTSAAPSESTPS 319
Db 251 SSNSGKEVLMPSHSLPPASLELSVTVKESPVLTFTPGSTEHSIPTPTSAAPSESTPS 310
Qy 320 ELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPAPPVETTYNVEWNLISHPT 379
Db 311 ELPISPTTAPRTVKELTVSAGDNLIIITLPDNEVELKAFVAPAPPVETTYNVEWNLISHPT 370
Qy 380 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSENAFGEFNVTVKPARRVNLPVAVVS 439
Db 371 DYQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSENAFGEFNVTVKPARRVNLPVAVVS 430
Qy 440 POLQELTLPLTSALIDGSGQSTDDETEIVSYHWEENGINEPIEEKTSVDSVPLRLSNLDPGNY 499
Db 431 POLQELTLPLTSALIDGSGQSTDDETEIVSYHWEENGINEPIEEKTSVDSVPLRLSNLDPGNY 490
Qy 500 SFRLTVTDSGATNSTTAALI VNNADVPPVANAGPNHTITLPQNSITLNGQSSDDHQI 559
Db 491 SFRLTVTDSGATNSTTAALI VNNADVPPVANAGPNHTITLPQNSITLNGQSSDDHQI 550
```

Qy 560 VLYEWSLGPSEGGKHVWVGQVOTPYLHLSAMQEGDYTFOLKVTDSRRQSTAXTVTVIOP 619  
 Db 551 VLYEWSLGPSEGGKHVWVGQVOTPYLHLSAMQEGDYTFOLKVTDSRRQSTAXTVTVIOP 610  
 Qy 620 ENNRPVAVAGPDKELIFFVESATLDGSSSSDDHGIYFVHWEHVRGSAVEMENIDKAI 679  
 Db 611 ENNRPVAVAGPDKELIFFVESATLDGSSSSDDHGIYFVHWEHVRGSAVEMENIDKAI 670  
 Qy 680 TVTGLQVGTYHFRUTVXKQGLSSTSTLTVAVKKENNSPPRARAGRHVLVLPNNISITLD 739  
 Db 671 TVTGLQVGTYHFRUTVXKQGLSSTSTLTVAVKKENNSPPRARAGRHVLVLPNNISITLD 730  
 Qy 740 GSRSTDORIVSYLWIRDOGSPPAGDVTDGSDHVALQTLNLVGVYTFHLRVTDGSGAS 799  
 Db 731 GSRSTDORIVSYLWIRDOGSPPAGDVTDGSDHVALQTLNLVGVYTFHLRVTDGSGAS 790  
 Qy 800 DTDATVEVQDPKRSGLVELTQVGVQOLTEQRKDTLVRQALVLLVDSIDKVKIRA 859  
 Db 791 DTDATVEVQDPKRSGLVELTQVGVQOLTEQRKDTLVRQALVLLVDSIDKVKIRA 850  
 Qy 860 HSDLTSTVVFYVQSRPPFKVLAABEARNLHMRLSKKADFLLPKVLKRVDTAGCLLKCSG 919  
 Db 851 HSDLTSTVVFYVQSRPPFKVLAABEARNLHMRLSKKADFLLPKVLKRVDTAGCLLKCSG 910  
 Qy 920 HGCDPLTKRCICSHLWENLIQRYINDGESNCEWSIFYVTVLAFTLVLVTGGFTWLCIC 979  
 Db 911 HGCDPLTKRCICSHLWENLIQRYINDGESNCEWSIFYVTVLAFTLVLVTGGFTWLCIC 970  
 Qy 980 CCKRQKTKRKTKTYILDNMDQERMLPKYGIKHSRTEHNSLVMSESEFSDQDT 1039  
 Db 971 CCKRQKTKRKTKTYILDNMDQERMLPKYGIKHSRTEHNSLVMSESEFSDQDT 1030  
 Qy 1040 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1072  
 Db 1031 IFSREKMERGNPKVSMNGSIRNGASFSYCSKDR 1063

RESULT 14

US-11-097-143-16884  
 ; Sequence 16884, Application US/11097143  
 ; Publication No. US20050208558A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Venter, J. Craig  
 ; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
 ; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
 ; TITLE OF INVENTION: DROSOPHILA GENES.  
 ; FILE REFERENCE: CL000728  
 ; CURRENT APPLICATION NUMBER: US/11/097,143  
 ; CURRENT FILING DATE: 2005-04-04  
 ; PRIOR APPLICATION NUMBER: 60/157,832  
 ; PRIOR FILING DATE: 1999-10-05  
 ; PRIOR APPLICATION NUMBER: 60/160,191  
 ; PRIOR FILING DATE: 1999-10-19  
 ; PRIOR APPLICATION NUMBER: 60/161,932  
 ; PRIOR FILING DATE: 1999-10-28  
 ; PRIOR APPLICATION NUMBER: 60/164,769  
 ; PRIOR FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: 60/173,383  
 ; PRIOR FILING DATE: 1999-12-28  
 ; PRIOR APPLICATION NUMBER: 60/175,693  
 ; PRIOR FILING DATE: 2000-01-12  
 ; PRIOR APPLICATION NUMBER: 60/184,831  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: 60/191,637  
 ; PRIOR FILING DATE: 2000-03-23  
 ; NUMBER OF SEQ ID NOS: 43008  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16884  
 ; LENGTH: 1069  
 ; TYPE: PRT  
 ; ORGANISM: DROSOPHILA

US-11-097-143-16884

Query Match 26.0%; Score 1451.5; DB 20; Length 1069;  
 Best Local Similarity 32.1%; Pred. No. 2,3e-80;  
 Matches 373; Conservative 172; Mismatches 383; Indels 233; Gaps 35;  
 Qy 8 LSSLLLVTTIAGCA-----RKCQSEGRYTSYNAISPMLTTRIMRVSHST 51  
 Db 8 ICNLLLLATAMSSAYADVTYQNALLVGSKKHETSPDNSVCGS-ISPMLVCHKMLR--HV 64  
 Qy 52 F-----PVVD-----CTAACCD-----LSSCDLAWWPEG 75  
 Db 65 FENATPRDEQAGVFEBYKPPDAVEPLEBEAYLWNCLOACCEKPRNGSSACNVVLFKA 124  
 Qy 76 RCVLVSPCHKENCBP--KMGPIRSYLTFLVRPV-----QRPALLDYGD 118  
 Db 125 KCVHRCQSNSEACLPKLRVM-PNEKVQVMLVNLPLGATWPLQLKAEAAKQNAEILPYDE 183  
 Qy 119 MMLNRGSPSGWGDSPEDIRKDLKFLCKDWGLEMSEYXDD-----VRELEKLLQPSGK 173  
 Db 184 AALN-----FWQP-----RRJSLARN---QETPVTEDEDFPLADKRMQMIFQDEN 229  
 Qy 174 QEPGRSAEYTDWGLLPGESEGAFFNSVGDSPAPVPAETQDPELHYNBSA-----STPA 226  
 Db 230 -----DVLANEELGYDYDSNAKFTTCDMETPC 255  
 Qy 227 PKUPERSVLLPLPTPSS-----GEVLEKSK-----ASQLOEBOSSNSSGKEV 268  
 Db 256 PP-POQCV-PLQNAVARGVCTCPGFFVWKNKORKCVMAAVPPYSSYL---TSNEAGQOE 307  
 Qy 269 LMPHSHLPASLELSSVTVVEKSPVLTVTPGSTSEHSIPTPTPSAAPSESTPSELPISTPA 328  
 Db 308 AAASENSPEVS-----TPPLKAEQN----- 327  
 Qy 329 PRTVKELTVSAGDNLIITLPDNEVELKAFVAPAPPV-ETTYNYEWNLSHPT-DYQGBIK 386  
 Db 328 ----KDIVSVMSK-EVRLPEQEVTLAAFTVPDQTSQTKYKYLWTLISQKGPWNGTIS 382  
 Qy 387 QGHQTLNLSQLSVGLYVFKVTVSSEN-AFEGGFVNTVTPARVNLPPVAVSPQLQEL 445  
 Db 383 DQSKSKVKLSNLSEGLYTFKVTVDNGTFCGATANTVLPENRINQPPQVVIIPREDII 442  
 Qy 446 TLPLTSALIDGQSTDDTEIVSVHWEIINGPFIKEKTSVDSVPLRLSNLDGNYSERLTV 505  
 Db 443 RQPTTNAILDGSTSTDDDKITNMHWEVIGSIPQVLPVNTLQDLTSPGNTYFKLTV 502  
 Qy 506 TDSGATNTTAAALIVNNAVDPYPVANAGNHTITLPONSIITLNGNOSDDHQVLYEWS 565  
 Db 503 TDSNVNTTATIAVLKETDYAPVANAGDAVILYLPNNVTLNGTASSDDHEIVAWEMT 562  
 Qy 566 LGPGSEKHHVWVGQVOTPYLHLSAMQEGDYTFOLKVTDSRRQSTAXTVTVIOPENRPP 625  
 Db 563 KDADEAKAVDMQNTRTPYVQLSNLEEGMTFVLKTDGSGQSTAKVHVFPVKPTNSPP 622  
 Qy 626 VAVAGPKELIFFVESATLDGSSSSDDHGIYFVHWEHVRGSAVEMENIDKAIATVGLQ 685  
 Db 623 VAEAGSNNTTSLPINWVLLNGSDKDDIGIKSYLWKQLSGPNNVILKSNSSIANATSLT 682  
 Qy 686 VGTVHFRUTVXKQGLSSTSTLTVAVKKENNSPPRARAGRHVLVLPNNISITLDGSRSTD 745  
 Db 683 LGLYEFELTVADENNTATDTTWVKIYQERNAAPINAGGDHTVTLPATAYFNGSKSWD 742  
 Qy 746 DQRIYSYLWIRDOGSPPAGDVTDGSDHVALQTLNLVGVYTFHLRVTDGSGASDTDTAT 805  
 Db 743 DLAVVKYLWTRDEHSLAAGVIVADTDKEPVMILNLVQGRVYFTLTVSDDDQGLTSSDFTVS 802  
 Qy 806 VEVPQDPKRSGLVELTQVGVQOLTEQRKDTLVRQALVLLVDSIDKVKQIR--AHSDL 863  
 Db 803 VNVRDEPKLNLVQMTLPMGIVSLVQSELDVSVQKQLQLLGD-ENKIQVRELKYLHPTD- 860  
 Qy 864 STVIVFYVQSRPPFKVLAABEARNLHMRLSKKADFLLPKVLKRVDTAGCLLKCSGHGHC 923  
 Db 861 ATVLVIFYNDGQG-KALDGLQVERQLRTQLQKQASILGAFV-DIRTSVCOSDCSGHSC 918

```

Qy 924 DPLPKRCICSHLWMENILQRYINDG--ESNCWESIFPYVTYVLAFTLIVLTGGFTWLCICCCCK 982
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 919 NPITRACICEAFWMPSP--AGYFFNNQANCDSILYVFGVIVGCLLLSGVFWGCIACACR 976
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Qy 983 RQKETKTRKKT-KYTILDNMDEQRMELRPKYGIGKHRSTEHNSSLWYSESEFSDSDQTIF 1041
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 977 QSKKPRLRQKQKYSLGNDKEAA-----NYSRNTSLTSESTSDVLF 1020
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Qy 1042 SREKMERGNPKVSMNGSIRNG 1062
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 1021 ETRTKSNGLGKHKSHNSHSG 1041
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

RESULT 15
US-10-450-763-39353
; Sequence 39353, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 39353
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(711)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-39353

```

```

764  QY 764  GDIVGSDHSVALQLNLNVEGVYTFHLRVDTDSOGASDTTATVVEVODPRKSGLVELTLQ 823
      Db      :
345  QY 345  QDIVGSDHSVALQLNLNVEGVYTFHLMXLSLCKCPKANPVSSLSFDPDRKSGLVELTLQ 404
      Db      :
824  QY 824  VGVGQLTEORQDPTLVRQLAVLNLVDSDIKVOKIRAHSDL-STVIVFYVQSRPEPKVLKA 882
      Db      :
405  QY 405  VGVGQLTEORQDPTLVRQLAVGVSWXGSPSDHV-AVLSHCLLHSTVIVFYVQSRPEPKVLKA 463
      Db      :
883  QY 883  AEVARNLHMLSKXEKADFLFKVLVRVDTAGCLLKCSGHGHCDELTKRCICSHLWMEN-LI 941
      Db      :
464  QY 464  AEVARNLHIXKSRDLSFCMLPSALHTSSAGCLLKCSGHGHCDELTKRCIXKXSWSTNSLX 523
      Db      :
942  QY 942  QRYIWGESNC--EWSIFYVTVLAFTLIVLTGFTWLCI-----CCCKROKRYKIRKKT- 993
      Db      :
524  QY 524  LCPLW---FSCWSEWSIFYVTVLKSLSV-----TFLKIXVELFIVFPFROKTKIRKKTLL 575
      Db      :
994  QY 994  -----KY-----TILD-----NMDEQEBWELRPKYGIKHRSTEHN 1023
      Db      :
576  QY 576  NMXXVHLUKYFTHXILPLIAVRGCMNDYPHDXQLEVNHGVPQOMLLCFSLGIKHRSTEHN 635
      Db      :
1024 QY 1024 SSLMVSESEFSDQDTIFS 1042
      Db      :
636  QY 636 SSLMVSESVSFDRDCLQS 654
      Db      :

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 12, 2005, 10:33:12 ; Search time 3809 Seconds  
(without alignments)  
1960.176 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSLLLVITAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table:

	BLOSUM62	Gapop 10.0	Xgapext 0.5
Gapop 10.0		0.5	
Xgapop 10.0			0.5
Gapop 6.0			7.0
Delop 6.0			7.0

Searched: 8443130 seqs, 3482420727 residues

Total number of hits satisfying chosen parameters: 16886260

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US10764390/runat\_12102005\_110204\_22383/app\_query.fasta\_1.1223  
-DB=published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10764390 @CGN 1 1 723 @runat\_12102005\_110204\_22383  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:\*

```

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5574	99.9	6791	20	US-10-764-390-2
2	5574	99.9	6791	20	US-10-764-390-4
3	5546.5	99.4	6797	20	US-10-764-390-270
4	5546.5	99.4	6797	20	US-10-764-390-271
5	5527	99.1	6991	20	US-10-764-390-6
6	5527	99.1	6991	20	US-10-764-390-269
7	1451.5	26.0	3345	24	US-11-097-143-16883
8	1314	22.5	2423	22	US-10-450-763-8985
9	1280.5	22.9	5775	24	US-11-097-143-16882
10	477	8.5	284	20	US-10-764-390-1
11	403	7.2	136990	19	US-10-717-597-68
12	397	7.1	1420	14	US-10-050-704-89
13	397	7.1	1420	19	US-10-798-512-89
14	241.5	4.3	3612	17	US-10-282-122A-13443
15	225	4.0	82027	21	US-10-278-698-1034
16	225	4.0	82027	21	US-10-278-698-1045
17	224	4.0	81940	9	US-09-759-508B-1
18	224	4.0	81940	10	US-09-360-706-1092
19	224	4.0	81940	10	US-09-873-319-724
20	224	4.0	81940	20	US-10-723-860-132
21	224	4.0	81940	20	US-10-656-873A-1
22	224	4.0	81940	22	US-10-756-149-144
23	221.5	4.0	110838	22	US-10-840-512-41
24	219.5	3.9	2379	21	US-10-875-518-3
25	214.5	3.8	3687	10	US-09-917-384-2
26	214.5	3.8	3687	10	US-09-917-383-2
27	214	3.8	23546	24	US-11-097-143-1223
28	213	3.8	16680	17	US-10-282-122A-38993
29	207	3.7	44014	24	US-11-097-143-1222
30	206.5	3.7	3586	17	US-10-282-122A-9979
31	204	3.7	8307	24	US-11-097-143-31982
32	204	3.7	11597	24	US-11-097-143-31981
33	202	3.6	2601	17	US-10-297-639-14
34	202	3.6	4951	22	US-10-450-763-23901
35	202	3.6	5403	21	US-10-956-157-1961
36	202	3.6	5403	22	US-10-450-763-13894
37	202	3.6	5403	22	US-10-450-763-29584
38	201.5	3.6	25165	14	US-10-114-170-39
39	201	3.6	3928	20	US-10-723-860-7627
40	199	3.6	2526	21	US-10-489-695-53
41	199	3.6	2537	21	US-10-489-695-52
42	199	3.6	17203	9	US-09-864-761-20867
43	199	3.6	17203	16	US-10-029-386-20814
44	199	3.6	182508	22	US-10-981-277-27
45	197	3.5	1555	19	US-10-470-390A-17

# ALIGNMENTS

## RESULT 1

US-10-764-390-2  
; Sequence 2, Application US/10764390  
; Publication No. US2004021421A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Jakobovits, Aya  
; APPLICANT: Challita-Eid, Pia M.  
; APPLICANT: Ge, Wangmao  
; APPLICANT: Faris, Mary  
; APPLICANT: Steven B. Kanner  
; APPLICANT: Juan J. Perez-Villar  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 254PD6B Useful in Treatment and Detection of  
; TITLE OF INVENTION: Cancer





Qy 261 SerAsnSerSerGlyIysGluValLeuMetProSerHisSerLeuProProAlaSerLeu 280  
Db 1292 AGCAACAGCTCTGAAAGAGGTTCATATGCTTCCCATAGTCTTCTCCGCGAAGCGCTG 1351  
Qy 281 GluLeuSerSerValThrValGluIysSerProValLeuThrValThrProGlySerThr 300  
Db 1352 GAGCTCAGTCAAGTCACCGTGAGAAAGCCAGTGCTCACAGTCAACCCCGGAGATCA 1411  
Qy 301 GluHisSerIleProThrProProThrSerAlaProSerGluSerThrProSerGlu 320  
Db 1412 GAGCACAGCATCCCAACACCTCCCACTAGCGAGCCCTCTGAGTCCACCCCATCTGAG 1471  
Qy 321 LeuProIleSerProThrThrAlaProArgThrValIysGluLeuThrValSerAlaGly 340  
Db 1472 CTACCCATATCTCTTACCCTCTCCAGGACAGTGAAGAACTTACGGTATCGCTGGA 1531  
Qy 341 AspAsnLeuIleIleThrLeuProAspAsnGluValGluLeuLysAlaPheValAlaPro 360  
Db 1532 GATAACCTAATTATACTTTACCCGACAACTGAAGTTGAAGTGAAGGCCCTTTGTTGGCCA 1591  
Qy 361 AlaProProValGluThrThrTyraSerThrGluThrProAsnLeuIleSerHisProThrAsp 380  
Db 1592 CGCGCACCTGTAGAAACAACCTACAACCTATGAATGAATTTAATAAGCCACCCACAGAC 1651  
Qy 381 TyrGlnGlyGluIleLysGlnGlyHisIysGlnThrLeuAsnLeuSerGlnSerVal 400  
Db 1652 TACCAGGTGAATTAATAACAAGGACACAAAGCAAACTCTTAACCTCTCTCAATTTGTCGTC 1711  
Qy 401 GlyLeuThrValPheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheVal 420  
Db 1712 GGACTTATGTTCTCAAGTCACTGTTTCTAGTGAAGCGCTTTGGAGAAGGATTTGTC 1771  
Qy 421 AsnValThrValLysProAlaArgValAsnLeuProProValAlaValIserPro 440  
Db 1772 AATGTCACTTAAAGCCTGCCAGAAGAGTCAACCTGCCACCTGTAGCAGTGTGTTCTCCC 1831  
Qy 441 GlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThr 460  
Db 1832 CAATGCAAGAGCTCACTTTGCCCTTGAAGTCAAGCGCTTCAATGAGGAGCAAGTACA 1891  
Qy 461 AspAspThrGluIleValSerThrHisThrGluLeuLeuAsnGlyProPheIleGluGlu 480  
Db 1892 GATGATCTGAATAGTGAATTCATTTGGAGAGAAATTAACGGGCGCTTTCATAGAGAG 1951  
Qy 481 LysThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnThrSer 500  
Db 1952 AAGACTTCAGTTGACTCTCCGCTCTACGTTGTCTAACCCTGTATCTGGTAACTATAGT 2011  
Qy 501 PheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrAlaAlaLeuIle 520  
Db 2012 TTCAGGTTGACTGTTACAGACTCGGAGCGGACCACTAATCTACAACTGCGAGCCCTAATA 2071  
Qy 521 ValAsnAsnAlaValAspThrProProValAlaAsnAlaGlyProAsnHisThrIleThr 540  
Db 2072 GTGAACAATGCTGGAGTACCCACGATGCTATATGAGGACCAATACACCAATAACT 2131  
Qy 541 LeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspHisGlnIleVal 560  
Db 2132 TTGCCCAAAATCCATCACTTGAATGMAAACACAGACAGTGAAGTCAACAGATGCTC 2191  
Qy 561 LeuThrGluThrProSerLeuGlyProGlySerGluGlyLysHisValValMetGlnGlyVal 580  
Db 2192 CTCTATGATGCTGCTCCGCTCGGAGTGAAGGCAACATGTGGTCAATGAGGAGTA 2251  
Qy 581 GlnThrProThrLeuHisLeuSerAlaMetGlnGluGlyAspThrThrPheGlnLeuLys 600  
Db 2252 CAGAGCCATACCTTCAATTTATCTGCAATGAGGAAGAGATATATACATTTTCAAGCTGAAG 2311  
Qy 601 ValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnProGlu 620  
Db 2312 GTGACAGATTTCTTCAAGCAACAGTCTACTGCTGTGTGACTGTGATTTGTCAGGCTGAA 2371  
Qy 621 AsnAsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGlu 640

Db 2372 AACATAGACTCCAGTGGCTGTGGCCGCGCTGATTAAGAGCTGATCTTCCAGTGGAA 2431  
Qy 641 SerAlaThrLeuAspGlySerSerSerSerSerAspHisGlyIleValPheThrHisTrp 660  
Db 2432 AGTGCTACCTCGATGGAGGAGCAGCAGCGCATGACCAACGCAATGTCTTCTACCACTGG 2491  
Qy 661 GluHisValArgGlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAlaThr 680  
Db 2492 GAGCACGTGAGAGCCCGCAGTGCAGTGGAGATGAAATAATTGACAAAGCAATAGCCACT 2551  
Qy 681 ValThrGlyLeuGlnValGlyThrTyraSerPheArgLeuThrValLysAspGlnGlnGly 700  
Db 2552 GTGACTGGTCTCCAGTGGGGACCTACCACTTCGTTTGACAGTGAAGACCAAGCAGGA 2611  
Qy 701 LeuSerSerThrSerThrLeuThrValAlaValLysLysGluAsnAsnSerProProArg 720  
Db 2612 CTGAGCAGCAGCTCCACCTCACTGTGGTGTGAAGAAAGAAATAATAGTCTCTCCACA 2671  
Qy 721 AlaArgAlaGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGly 740  
Db 2672 GCGCGGCTGTGGCAGACATGTTCTTGTGCTTCCCAATAATTCATTTTGGATGGT 2731  
Qy 741 SerArgSerThrAspAspGlnArgIleValSerTyrLeuTrpIleArgAspGlyGlnSer 760  
Db 2732 TCAAGTCTACTGATGACCAAGAAATGTGCTCTATCTGTGGATCCGGATGGCCAGAGT 2791  
Qy 761 ProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsn 780  
Db 2792 CCAGCAGCTGGAGATGATCATCGATGGCTCTGACCACAGTGTGGCTCTGCAGCTTACGAAT 2851  
Qy 781 LeuValGluGlyValThrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAsp 800  
Db 2852 CTGGTGGAGGGGTGTACACTTCCACTTCCGAGTCAACGACAGTCAAGGGGGCTCGGAC 2911  
Qy 801 ThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValGluLeu 820  
Db 2912 ACAGACACTGCACTGTGGAAGTGCAGCCAGCCCTAGGAAGAGTGGCTCTGGTGGAGCTG 2971  
Qy 821 ThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAspThrLeuValArgGln 840  
Db 2972 ACCCTGCAGGTGGTGTGGCAGCTGACAGCAGCAGCGGAAGGACACCTTGTGAGGCAG 3031  
Qy 841 LeuAlaValLeuLeuAsnValLeuAspSerAspIleLysValGlnLysIleArgAlaHis 860  
Db 3032 CTGGCTGTGCTGTGAACGCTGCGACTCGGACATTAAGGTCCAGAAAGATTCGGGCCAC 3091  
Qy 861 SerAspLeuSerThrValIleValPheThrValGlnSerArgProProPheLysValLeu 880  
Db 3092 TCGGATCTCAGCACCGTGAATGTGTTTATGTACAGAGCAGCGCCCTTCAAGGTTCTC 3151  
Qy 881 LysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAspPhe 900  
Db 3152 AAAGCTGCTGAAGTGGCCCGAAATCTGCACATGCGCTCTCAAAGGAGAGGCTGACTTC 3211  
Qy 901 LeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHis 920  
Db 3212 TTGCTTTTCAAGTCTTGAAGGTTGATACAGCAGGTTCCTTCTGAAGTGTCTTGGCCAT 3271  
Qy 921 GlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAsnLeu 940  
Db 3272 GGTCACTCGGACCCCTCACAAGCGCTGCAATTTGCTCTCACTTATGATGAGAACCTT 3331  
Qy 941 IleGlnArgTyrIleTrpAspGlyGluSerAsnCysGluTrpSerIlePheThrValThr 960  
Db 3332 ATACAGGTTATATCTGGATGGAGAGCAACTGTGAGTGGAGTATATTCTATGTGACA 3391  
Qy 961 ValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCys 980  
Db 3392 GTGTTGGCTTTTACTTATTTGTCTAAGCAGAGGTTTCACTTGGCTTGTGATCTGTGTC 3451  
Qy 981 CysLysArgGlnLysArgThrLysIleArgLysLysThrLysTyrThrIleLeuAspAsn 1000

Db 3452 TCACAAAGACAAAAAGGACTAAAAATCAGGAAAAAACAAGTACACCATCTCTGGATAAC 3511  
QY 1001 MetAspGluGlnArgMetGluLeuArgProLysTyrGlyIleLysHisArgSerThr 1020  
Db 3512 ATGGATGAACAGAAAAAGATGAAGTGGAGCCCAATATGGTATCAGACCCGAAGCACA 3571  
QY 1021 GluHisAsnSerSerIleuMetValSerGluSerGluPheAspSerAspGlnAspThrIle 1040  
Db 3572 GAGCACAACTCCAGCCTGATGTATCCGAGTCTGAGTTTGACAGTGCACAGGACACAATC 3631  
QY 1041 PheSerArgGluLysMetGluArgGlyAsnProLysValSerMetAsnGlySerIleArg 1060  
Db 3632 TTCAGCCGAGAAAAAGATGAGAGAGAGGAATCCAAAAGGTTTCATGAATGGTTCCATCAGA 3691  
QY 1061 AsnGlyAlaSerPheSerTyrCysSerLysAspArg 1072  
Db 3692 AATGAGAGCTTCCTTCAGTTATTGCTCAAAAGGACAGA 3727

## RESULT 2

US-10-764-390-4  
; Sequence 4, Application US/10764390  
; Publication No. US2004021412A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Jakobovits, Ava  
; APPLICANT: Challita-Bid, Pia M.  
; APPLICANT: Ge, Wangmao  
; APPLICANT: Faris, Mary  
; APPLICANT: Steven B. Kanner  
; APPLICANT: Juan J. Perez-Villar  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 254P1D6B Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20081.00  
; CURRENT APPLICATION NUMBER: US/10/764,390  
; CURRENT FILING DATE: 2004-01-23  
; PRIOR APPLICATION NUMBER: US60/442,526  
; PRIOR FILING DATE: 2003-01-24  
; NUMBER OF SEQ ID NOS: 277  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 6791  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (512)... (3730)  
US-10-764-390-4

Alignment Scores:  
Pred. No.: 0 Length: 6791  
Score: 5574.00 Matches: 1069  
Percent Similarity: 99.72% Conservative: 0  
Best Local Similarity: 99.72% Mismatches: 3  
Query Match: 99.89% Indels: 0  
DB: 20 Gaps: 0

US-10-764-390-3 (1-1072) x US-10-764-390-4 (1-6791)

QY 1 MetAlaProProThrGlyValLeuSerSerLeuLeuLeuValThrIleAlaGlyCys 20  
Db 512 ATGGCGCCCCCACAGGTGTGCTCTCTTCATTGCTGCTGCTGTTGATGCAATTCAGGTGT 571  
QY 21 AlaArgLysGlnCysSerGluArgThrTyrSerAsnAlaValIleSerProAsnLeu 40  
Db 572 GCCCGTAAGCAGTGCAGCGAGGGGAGACATATTCCAATGCAGTCATTTACCTAACTTG 631  
QY 41 GluThrThrArgIleMetArgValSerHisThrPheProValValAspCysThrAla 60  
Db 632 GAAACCCAGCAATCATCGGGGTCTCACACCTTCCCTGCTGATAGACTGCACGGCGCT 691  
QY 61 CysCysAspLeuSerSerCysAspLeuAlaTrpTrpPheGluGlyArgCysTyrLeuVal 80

Db 692 TGCCTGACCTGTCCAGCTGTGACCTGGCTGGTTCGAGGGCCGCTGCTACCTGGTGTG 751  
QY 81 SerCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTyrLeu 100  
Db 752 AGCTGCCCCCAACAAGAGAACTGTGAGCCCAAGAGATGGGCCCATCAGGTCTTATCTC 811  
QY 101 ThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMet 120  
Db 812 ACTTTTGTGCTCCGGCCTGTTTCAGAGGCTGCACAGCTGCTGGACTATGGGGACATGATG 871  
QY 121 LeuAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArgLysAsp 140  
Db 872 CTGAACAGGGGCTCCCTCCGGGATCTGGGGGACTCACCTGAGGATATCAGAAAGGAC 931  
QY 141 Leu\*\*\*PheLeuGlyLysAspTrpGlyLeuGluMetSerGlyTyr\*\*\*AspAspTyr 160  
Db 932 TTGCCCTTTTAGGCAAGATTTGGGGCTAGAGAGATGTCTGAGTACGCAGATACACTAC 991  
QY 161 ArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAla 180  
Db 992 CGGGAGCTGGAGAAGGACCTCTTGCNACCCAGTGGCAAGCAGGAGCCACAGAGGAGTGCC 1051  
QY 181 GluTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGly 200  
Db 1052 GAGTACACGGACTGGGGCTACTGCGGGCAGCGAGGGGGCTTCAACTCTCTCTGTGGA 1111  
QY 201 AspSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGlu 220  
Db 1112 GACAGTCTCGGGTGCCACAGCGAGACCGCAGCAGACCTTGAGCTCCATTACCTGAATGAG 1171  
QY 221 SerAlaSerThrProAlaProLysLeuProGluArgSerValLeuLeuProLeuProThr 240  
Db 1172 TCGGCTTCAACCCCTGCCCCCAAACTCCCTGAGAGAGTGTGTGCTTCCCTTGGCGACT 1231  
QY 241 ThrProSerSerGlyGluValLeuGluLysGluLysAlaSerGlnLeuGlnGlnSer 260  
Db 1232 ACTCCATCTTCAGGAGAGGTGTGAGAGAAAGAAAGGCTTCTCAGCTCCAGGAACAATCC 1291  
QY 261 SerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProProAlaSerLeu 280  
Db 1292 AGCAACAGCTCTGAAAGAGAGGTCTAATGCTTCCCATAGTCTTCTCCGGCAAGCCTG 1351  
QY 281 GluLeuSerSerValThrValGluLysSerProValLeuThrValThrProGlySerThr 300  
Db 1352 GAGCTCAGCTCAGTCACCGTGGAGAAAGCCAGTGCTCACAGTCACCCCGGGGAGTACA 1411  
QY 301 GluHisSerIleProThrProThrSerAlaAlaProSerGluSerThrProSerGlu 320  
Db 1412 GAGCACAGCATCCCAACACCTCCCACTAGCGAGCGCCCTCTGAGTCCACCCCATCTGAG 1471  
QY 321 LeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAlaGly 340  
Db 1472 CTACCCCATATCTCTTACCACCTGCTCCAGGACAGTGAAGAACTTACCGGTATCGGCTGA 1531  
QY 341 AspAsnLeuIleIleThrLeuProAspAsnGluValGluLeuLysAlaPheValAlaPro 360  
Db 1532 GATAACCTTAATTATTAACCTTACCCGACAAATGAAGTTGAATGAAGCCCTTTGTTGGCCA 1591  
QY 361 AlaProProValGluThrThrTyrAsnTyrGluTrpAsnLeuIleSerHisProThrAsp 380  
Db 1592 GCGCCACCTGTAGAAAACAACCTTACAACTATGAATGGAATTTAATAAGCCACCCACAGAC 1651  
QY 381 TyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerVal 400  
Db 1652 TACCAGGTGAAATAAAAACAAGGACACAGCAAACTCTTAACCTCTCTCAATTGTCGGTC 1711  
QY 401 GlyLeuTyrValPheLysValThrValSerSerGluAsnAlaPheGlyGlyPheVal 420  
Db 1712 GGACTTTATGCTTCAAAGTCACTGTTTCTAGTGAAGACGCTTTGGAGAGGATTTGTC 1771  
QY 421 AsnValThrValLysProAlaArgValAsnLeuProProValAlaValValSerPro 440



1234	GACTACTCCATCTTTCAGGAGAGGTGTTGGAGAAAGAAAAGGCTTCTTCAGCTCCAGAAC	1239	
Qy	259	nSerSerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProProlase	279
Db	1294	ATCCAGCAACAGCTCTGGAAAGAGGTTCTAATGCCTTCCCATAGTCTTCTCCGCAAG	1353
Qy	279	rLeuGluLeuSerSerValThrValGluLysSerProValLeuThrValThrProGlySe	299
Db	1354	CCTGGAGCTCAGCTCAGTCACCGTGAGAAAGCCAGTGCTCACAGTCACCCCGGGAG	1413
Qy	299	rThrGluHisSerIleProThrProProThrSerAlaAlaProSerGluSerThrProSe	319
Db	1414	TACAGAGCAGCATCCCAACACTCCCACTAGCGGAGCCCTCTGAGTCCACCCCATC	1473
Qy	319	rGluLeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAl	339
Db	1474	TGAGCTACCCATATCTCTACCACTGCTCCAGGACAGTGAAGAACTTACGGTATCGGC	1533
Qy	339	aGlyAspAsnLeuIleIleThrLeuProAspIenGluValGluLeuLysAlaPheValAl	359
Db	1534	TGAGAGTAACCTAAATTATTAACCTTACCCGACATGAAGTTGAACCTGAAGGCCCTT	1593
Qy	359	aProAlaProProValGluThrThrTrpAsnTrpGluTrpAsnLeuIleSerHisProTh	379
Db	1594	GCAGCGCCACCTGTGAGAACAACTTACAACATATGAATGGAATTTAATAGCCACCCAC	1653
Qy	379	rAspTrpGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSe	399
Db	1654	AGACTACCAAGTGAATAAACAAGGACACACAGCAAACTCTTAACCTCTCTCAATTGTC	1713
Qy	399	rValGlyLeuTrpValPheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPh	419
Db	1714	CGTCGGACTTTATGTCTTCAAGTCACTGTTCTAGTGAAAACGCCCTTTGGAGAAGGATT	1773
Qy	419	eValAsnValThrValLysProAlaArgValAsnLeuProProValAlaValLysSe	439
Db	1774	TGTCAAATGTCACTGTTAAGCCTGCCAGAGAGTCAACCTGCCACCTGTAGCAGTGTGTT	1833
Qy	439	rProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSe	459
Db	1834	TCCCCAACTGCAAGAGCTCACTTGTGCTTTGAGCTCAGCCCTCAATTGATGGCAGCCNAAG	1893
Qy	459	rThrAspAspThrGluIleValSerTrpHisStrpGluGluIleAsnGlyProPheIleGl	479
Db	1894	TACAGATGATACTGAAATAGTGAGTTATCATTTGGGAGAGAAATAACGGGCCCTTCATAGA	1953
Qy	479	uGluLysThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTy	499
Db	1954	AGAGAAAGACTTCAGTTGACTCTCCCGCTCTTACCGCTTGTCTAAACCTTGATCCTCGTAACTA	2013
Qy	499	rSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLe	519
Db	2014	TAGTTTCAGGTGACTGTTACAGACTCGGACGGAGCCACTAACTCTAACACTGCAGGCCCT	2073
Qy	519	uIleValAsnAsnAlaValAspTrpProProValAlaAsnAlaGlyProAsnHisThrIle	539
Db	2074	AATAGTGAAACAATGCTGTGACTACCACCCAGTTGCTAATGCAGGACCAAAATCACACCAT	2133
Qy	539	eThrLeuProGlnAsnSerIleThrIleuIenGlyAsnGlnSerSerAspAspHisGlnIle	559
Db	2134	AACTTTGCCCAAAAACCTCCCATCACTTTGAATGAAACACAGAGAGAGTACCATCACCAAGAT	2193
Qy	559	eValLeuTrpGluTrpSerLeuGlyProGlySerGluGlyLysHisValValMetGlnGly	579
Db	2194	TGTCCTCTATGAGTGTGTCCTCGGGCTCTGGAGGTGAGGGCAACAATGTGTCTATGCAAGG	2253
Qy	579	yValGlnThrProTrpTrpLeuHisLeuSerAlaMetGlnGlnGlyAspTrpThrPheGlnLe	599
Db	2254	AGTACAGACCCCATACCTTCATTTAATCTGCAATGCAGGAAGGAGATTATACATTTTCAGCT	2313
Qy	599	uLysValThrAspSerSerArgGlnGlnSerThrAla***ValThrValIleValGlnPr	619
Db	2314	GAAAGTCACAGATTCTTCAAGGCAACAGTCTACTGCTGTGGTGACTGTGATGTGTCAGCC	2373

```
Qy 619 oGluAnsnAnArgProProValAlaValAlaGlyProAspLysGluLeuLlePheProVa 639
Db 2374 TGAACAACATAGACCTCCAGTGGCTGTGGCGGCCCTGTATAAGAGAGCTGTATCTCCCACT 2433
Qy 639 lGluSerAlaThrLeuAspGlySerSerSerSerSerSerSerSerSerSerSerSerSer 659
Db 2434 GGAAAGTGTACCTCCCTGGATGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2493
Qy 659 sTriPgluHisValArgGlyProSerAlaValAlaGluMetGluAnleAspLysAlaIleAl 679
Db 2494 CTGGGAGCACGTCAGAGGCCCCAGTGCAGTGGAGATGGAATAATATGCAAGCAATAGC 2553
Qy 679 aThrValThrGlyLeuGlnValGlyThrTyxHisPheArgLeuThrValLysAspGlnGl 699
Db 2554 CACTGTGACTGCTCTCCAGGTGGGACCTACCCACTTCCGTTTGACAGTGAAGACCAAGCA 2613
Qy 699 nGlyLeuSerSerThrSerThrLeuThrValAlaValLysLysGluAnsnAnsnSerProPr 719
Db 2614 GGGACTGAGCAGCAGCTCACCCCTCACTGTGGCTGTGAAGAGGAAATAATATAGTCCTCC 2673
Qy 719 oArgAlaArgAlaGlyArgHisValLeuValLeuProAnsnAnsnSerLleThrLeuAs 739
Db 2674 CAGAGCCCGGCTGTGGCAGACATGTTCTTGTGCTTCCCAATAATTCATTTCTTGA 2733
Qy 739 pGlySerArgSerThrAspAspGlnArgIleValSerTyxLeuTrpLleArgAspGlyGl 759
Db 2734 TGGTTCAGAGGTCTACTGTATGACCAAGAAATGTGTCTTATCTGTGGATCCGGGATGGCCA 2793
Qy 759 nSerProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuTh 779
Db 2794 GAGTCCAGCAGCTGGAGATGTCATCGATGCTGTGACCAAGATGGCTGTGAGCTTAC 2853
Qy 779 rAsnLeuValGluGlyValTyxThrPheHisLeuArgValThrAspSerGlnGlyAlaSe 799
Db 2854 GAACTGGTGGAGGGGTGTACACTTTCCACTTGGAGTCCAGCAGCTCAGGGGGCTC 2913
Qy 799 rAspThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValGl 819
Db 2914 GGACACAGACACTGCACTGTGGAGTGCAGCCAGACCTAGGAAGAGTGGCTGTGGTGA 2973
Qy 819 uLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAspThrLeuValAr 839
Db 2974 GCTGACCTCGAGGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3033
Qy 839 gGlnLeuAlaValLeuLeuAnsnValLeuAspSerAspLleLysValGlnLysIleArgAl 859
Db 3034 GCAGCTGGCTGTGCTGACGTGCTGCACTGGACATTAAGGTCCAGAGATTCGGGC 3093
Qy 859 aHisSerAspLeuSerThrValIleValPheTyxValGlnSerArgProPheLysVa 879
Db 3094 CCACTCGGATCTCAGCACCGTGTATGTTTATGTATGATACAGAGCGCGCTTCAAGGT 3153
Qy 879 lLeuLysAlaAlaGluValAlaAArgAsnLeuHisMetArgLeuSerLysGluLysAlaAs 899
Db 3154 TCTCAAGCTGCTGAAGTGGCCCGGAAATCTGCACATGGCGTCTCAAGAGGAGGCTGA 3213
Qy 899 pPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGl 919
Db 3214 CTTCTGTGTTTCAAGGTCTTGAGGGTGTATACAGCAGTGGCTTCTGAAGTGTCTGG 3273
Qy 919 yHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAs 939
Db 3274 CCATGGTCACTGCGAGCCCTCACAAAGCGGTGCAATTTGCTCTCACTATGATGGAGAA 3333
Qy 939 nLeuIleGlnArgTyxIleTrpAspGlyGluSerAnsnCysGluTrpSerLlePheTyxVa 959
Db 3334 CCTTATACAGCGTTATATCTGGATGGAGAGCAACTGTGATGGAGATATATCTATGT 3393
Qy 959 lThrValLeuAlaPheThrLeuLleValLeuThrGlyLysPheThrTrpLeuCysIleCy 979
Db 3394 GACAGTGTGGCTTACTTATTTGCTTAACAGAGGTTTCACTTGGCTTGTCACTG 3453
```

```
Qy 979 sCysCysLysArgGlnLysArgThrLysIleAArgLysThrLysTyxThrLleLeuAs 999
Db 3454 CTGCTGCAAAAGACAAAAAAAGGACTAAATCAGGAAAAAAACAAAGTACACCATCTTGA 3513
Qy 999 pAsnMetAspGluGlnGluArgMetGluLeuArgProLysTyxGlyLleLysHisArgSe 1019
Db 3514 TAACATGGATGAACAGAGAAAGATGGAACCTGAGGCCCAATATGTTATCAAGCAGCCGAAG 3573
Qy 1019 rThrGluHisAnsnSerSerLeuMetValSerGluSerGluPheAspSerAspGlnAspTh 1039
Db 3574 CACAGACCAACTCCAGCCTGATGGTATCCGAGTCTGAGTTTGACAGTGCAGGACAC 3633
Qy 1039 rLlePheSerArgGluLysMetGluArgGlyAsnProLysValSerMetAnsnGlySerIl 1059
Db 3634 AATCTTCAGCCGAGAAAGATGGAGAGGAAATCCAAAGGTTTCCATGAATGGTTCCAT 3693
Qy 1059 eArgAsnGlyAlaSerPheSerTyxSerSerLysAspArg 1072
Db 3694 CAGAAATGGAGCTTCTTCTCAGTTATTTGCTCAAAGGACAGA 3733

RESULT 4
US-10-764-390-271
; Sequence 271, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 271
; LENGTH: 6797
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-764-390-271

Alignment Scores:
Pred. No.: 0 Length: 6797
Score: 5546.50 Matches: 1069
Percent Similarity: 99.53% Conservative: 0
Best Local Similarity: 99.53% Mismatches: 3
Query Match: 99.40% Indels: 2
DB: 20 Gaps: 1

US-10-764-390-3 (1-1072) x US-10-764-390-271 (1-6797)

Qy 1 MetAlaProThrGlyValLeuSerSerLeuLeuLeuValThrIleAla---Gl 19
Db 514 ATGGCGCCCCCACAGGTGTCTCTTCTATGCTGTGTGGACAAATTCAGTTTGG 573
Qy 19 YCysAlaArgLysGlnCysSerGluGlyArgThrTyxSerAsnAlaValIleSerProAs 39
Db 574 TTGTGCCCGTAAGCAGTGCAGCGAGGAGGAGCATATTCATGAGTCATTTACCTAA 633
Qy 39 nLeuGluThrThrArgIleMetArgValSerHisThrPheProValValAspCysThrAl 59
Db 634 CTTGGAAACCAACCAAGAAATCATGCGGTGTCTCACACCTTCCCTGCTAGACTGCACGCG 693
Qy 59 aAlaCysCysAspLeuSerSerCysAspLeuAlaTrpPheGluGlyArgCysTyxLe 79
```



Db 694 CGCTTGTGTGACCTGTCCAGCTGTGACCTGGCTGGTGGTTCGAGGGCCGCTGTACCT 753  
Qy 79 uValSerCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTy 99  
Db 754 GGTGAGCTGCCCCCAAGAGAACTGTGAGCCCAAGAGATGGGCCCCATCAGTCTTA 813  
Qy 99 rLeuThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspWe 119  
Db 814 TCTCACTTTTGTCTCCGSCCTGTTTCAGAGGCTCTGCACAGCTGTCTGGACTATGGGACAT 873  
Qy 119 tMetLeuAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArgLys 139  
Db 874 CATGCTGAACAGGGGCTCCCCCTCGGGATCTGGGGGACTCACCTGAGGATATCAGAAA 933  
Qy 139 sAspLeu\*\*\*PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTyr\*\*\*AspAs 159  
Db 934 GGACTTGGCCCTTCTTAGGGCAAGATTGGGGCTTAGAGGAGATGTCTGAGTACTCAGATGA 993  
Qy 159 pTyrArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlyse 179  
Db 994 CTACCGGGAGCTGGAGAGGACCTCTTGCACCCAGTGGCAAGCAGGAGGCCAGAGGGAG 1053  
Qy 179 rAlaGluTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerVa 199  
Db 1054 TGCCGAGTACACGGACTGGGGCTTACTGCCGGGCGAGCGAGGGGGCTTCAACTCCTCTGT 1113  
Qy 199 lGlyAspSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAs 219  
Db 1114 TGGAGACAGTCTCGCGTGCCAGCGGAGCGGACGACAGGACCCTGAGCTCCATTACTGAA 1173  
Qy 219 nGluSerAlaSerThrProAlaProLysLeuProGluArgSerValLeuLeuProLeuPr 239  
Db 1174 TGAGTGGCTTCAACCCCTGCCCAAACTCCCTGAGAGAGTGTGTCTTCCCTTGC 1233  
Qy 239 oThrThrProSerSerGlyGluValLeuGluLysGluLysAlaSerGlnLeuGlnGlu 259  
Db 1234 GACTACTCCATCTTCAGGAGAGGTGTGGAGAAAGAAAGGCTTCTCAGCTCCAGGAA 1293  
Qy 259 nSerSerAsnSerGlyLysGluValLeuMetProSerHisSerLeuProProAlaSe 279  
Db 1294 ATCCAGCAACAGCTCTGGAAAAGAGGTCTTAATGCTTCCCANAGTCTTCTCCGGCAAG 1353  
Qy 279 rLeuGluLeuSerSerValThrValGluLysSerProValLeuThrValThrProGlyse 299  
Db 1354 CTTGGAGCTCAGTCACTCCGTGGAGAAAGCCAGTGTCTCAGACTACCCCGGGAG 1413  
Qy 299 rThrGluHisSerIleProThrProProThrSerAlaProSerGluSerThrProSe 319  
Db 1414 TACAGAGCACAGCATCCCCAACACCTCCCACTAGCGCAGCCCCCTCTGAGTCCACCCCATC 1473  
Qy 319 rGluLeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAl 339  
Db 1474 TGAGCTTACCATATCTCTTACCCTCTCCAGGACAGTGAAGAACTTACCGGTATCGGC 1533  
Qy 339 aGlyAspAsnLeuIleIleThrLeuProAspAsnGluValGluLeuLysAlaPheValAl 359  
Db 1534 TGGAGATAACCTAAATTAATCTTACCCGACAATGAAGTTGAATGAAGGCTTTGTTC 1593  
Qy 359 aProAlaProProValGluThrThrTyrAsnTyrGluTrpAsnLeuIleSerHisProTh 379  
Db 1594 GCCAGCGCCACCTGTAGAAACCAACTACAACTATGAATGGAATTTAATAAGCCACCCAC 1653  
Qy 379 rAspTyrGlnGlyLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSe 399  
Db 1654 AGACTACCAAGGTGAATTAACACAGGACACAGCAAACTCTTAACCTCTCTCAATGTTC 1713  
Qy 399 rValGlyLeuTyrValPheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPh 419  
Db 1714 CGTCGGACTTTATGTCTTCAAGTCACTGTCTTCTAGTGAACCGCCCTTTGGAGAGGATT 1773  
Qy 419 eValAsnValThrValLysProAlaArgValAsnLeuProProValAlaValValSe 439  
Db 1774 TGTCAATGTCACTGTTAAGCCTGCCAGAAAGAGTCAACCTGCCACCTGTAGCAGTGTGTTTC 1833

Qy 439 rProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSe 459  
Db 1834 TCCCAACTGCAAGAGCTCACTTTGCCCTTGACGCTAGCCCTCATTTGATGCAGCAAG 1893  
Qy 459 rThrAspAspThrGluLysValSerTyrHisTrpGluGluLeuAsnGlyProPheIleGl 479  
Db 1894 TACAGATGATACTGAAATAGTGAATTATCATTTGGGAAGAAATAAACGGGCCCTTCATAGA 1953  
Qy 479 uGluLysThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTy 499  
Db 1954 AGAAGAGACTTCAGTTGACTCTCCCGCTTACCGCTTGTCTAACCTTGTATCTGTGTAATA 2013  
Qy 499 rSerPheArgLeuThrValThrAspSerAspGlyValAlaThrAsnSerThrThrAlaAla 519  
Db 2014 TAGTTTCAGTTGACTGTTACAGACTCGGACGGAGCCACTAACTCTCAACTGCAGCCCT 2073  
Qy 519 uIleValAsnAsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIle 539  
Db 2074 AATAGTGAACAATGCTGTGACTACCCACCACTGGTAAATGCAGGACCAAAATCACACCAT 2133  
Qy 539 eThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIle 559  
Db 2134 AACTTTGCCCAAACTCCATCACTTTGAATGGAAACCGAGACGAGTACGATCACAGAT 2193  
Qy 559 eValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyLysHisValValMetGlnGl 579  
Db 2194 TGTCTCTATGAGTGGTCCCTGGGTCTCTGGGAGTGAGGGCAACATGTGGTCATCAGGG 2253  
Qy 579 yValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLe 599  
Db 2254 AGTACAGCGCATACCTTCACTTTATCTGCAATGCAGGAGGAGATTATACATTTTCAGCT 2313  
Qy 599 uLysValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnPr 619  
Db 2314 GAAGTGACATCTTCAAGGCAACAGTCTACTGCTGTGTGAGCTGTGATTTGTCCAGCC 2373  
Qy 619 oGluAsnAsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProVa 639  
Db 2374 TGAAACANATAGACTCCAGTGGCTGTGGCCGGCCCTGATTAAGAGCTGATCTTCCAGT 2433  
Qy 639 lGluSerAlaThrLeuAspGlySerSerSerSerAspAspHisGlyIleValPheTyrHi 659  
Db 2434 GGAAAGTCTACTACCTGGATGGAGCAGCAGCAGCGCATCACCACGGCATTTGTCTTACCA 2493  
Qy 659 sTrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAl 679  
Db 2494 CTGGAGCACGTTCAGAGGCCCCAGTGCAGTGGAGATGGAAATATTGACAAAGCAATAGC 2553  
Qy 679 aThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrValLysAspGlnGl 699  
Db 2554 CACTGTGACTGTCTCCAGGTGGGACCTACCACCTTCCGTTTGACAGTGAAGACCA 2613  
Qy 699 nGlyLeuSerSerThrSerThrLeuThrValAlaValLysLysGluAsnAsnSerProPr 719  
Db 2614 GGGACTGAGCAGCAGCTCCACCTCACTGTGGCTGTGAAGAGGAAAAATAATAGTCTCTCC 2673  
Qy 719 oArgAlaArgAlaGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAs 739  
Db 2674 CAGAGCCCGGGTGTGGCAGACATGTTCTGTGTCTTCCCAATAATTCCATTACTTGA 2733  
Qy 739 pGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuTrpIleArgAspGlyGl 759  
Db 2734 TGGTTCAAGTCTACTGATGACCAAGAAATTTGTCTTCTGATCTCGGATGGCCA 2793  
Qy 759 nSerProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuTh 779  
Db 2794 GAGTCCAGCAGCTGGAGATGTCTCGATGGCTCTGACCACAGTGTGGCTCTCAGCTTAC 2853  
Qy 779 rAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSe 799  
Db 2854 GAATCTGGTGGAGGGGTGTACACTTTCCACTTTGCGAGTCCCGACAGTCAGGGGGCCTC 2913



QY 799 rAspThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValG1 819  
 Db 2914 GGACACAGACTGCCACTGTGGAGTGCAGCCAGACCTTAGGAGAGTGGCCCTGGTGA 2973  
 QY 819 uLeuThrLeuGlnValGlyValGlnLeuThrGluGlnArgLysAspThrLeuValAr 839  
 Db 2974 GCTGACCTGTCAGGTTGGTGGCAGCTCAGAGAGCAGCGAAGACACCCCTGTGAG 3033  
 QY 839 gGlnLeuAlaValLeuLeuAnValLeuAspSerAspIleLysValGlnLysIleArgAl 859  
 Db 3034 GCAGCTGCTGCTGCTGCTGAACCTGCTGAGCTCGGACATTAAGGTCCAGAAGATTGGGC 3093  
 QY 859 aHisSerAspLeuSerThrValIleValPheTyrValGlnSerArgProProPheLysVa 879  
 Db 3094 CCATCGGATCTCAGCACCGGTGATGTTTATGTATGATGAGAGCGCCCTTTCAAGGT 3153  
 QY 879 lLeuLysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAs 899  
 Db 3154 TCTCAAGCTGCTGAAGTGGCCCGGAATCTGCACATGCGGCTCTCAAGAGGAGAGGCTGA 3213  
 QY 899 pPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerG1 919  
 Db 3214 CTCTCTGCTTTTCAAGGCTTTCAGGGTTGATACAGCAGGTTGCCCTTCTGAAAGTGTCTGG 3273  
 QY 919 yHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTTPMetGluAs 939  
 Db 3274 CCATGCTCAGTGCAGACCCCTCACAAAGCGCTGCATTTGCTCTCACTTATGATGAGGAA 3333  
 QY 939 nLeuIleGlnArgTyrIleTTPAspGlyGluSerAsnCysGluTTPSerIlePheTyrVa 959  
 Db 3334 CCTTATACAGGTTATATCTGGATGGAGAGCAACTGTGAGTGGATATATCTATGT 3393  
 QY 959 lThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTTPLeuCysIleCy 979  
 Db 3394 GACAGTGTGGCTTTTACTCTTATTTGCTAACAGGAGGTTTCACTTGGCTTTGTCATCTG 3453  
 QY 979 sCysCysLysArgGlnLysArgThrLysIleArgLysLysThrLysThrIleLeuAs 999  
 Db 3454 CTGCTGCAAAAGACAAAAGACTTAAATCAGGAAAAAACAAGTACACCATCTCTGGA 3513  
 QY 999 pAsnMetAspGluGlnGluArgMetGluLeuArgProLysTyrGlyIleLysHisArgSe 1019  
 Db 3514 TAACATGATGAACAGGAAGATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 3573  
 QY 1019 rThrGluHisAsnSerLeuMetValSerGluSerGluPheAspSerAspGlnAspTh 1039  
 Db 3574 CACAGACACAACTCCAGCTGATGGTATCCGAGTCTGAGTTTACAGTGCAGCAGACAC 3633  
 QY 1039 rIlePheSerArgGluLysMetGluArgGlyAsnProLysValSerMetAsnGlySerIl 1059  
 Db 3634 AATCTTCAGCCGAGAAAAGATGGAGAGGGAATCCAAAGGTTTCCATGAATGGTTCCAT 3693  
 QY 1059 eArgAsnGlyAlaSerPheSerTyrCysSerLysAspArg 1072  
 Db 3694 CAGAAATGGAGCTTCTTCAGTTATTTGCTCAAAAGGACAGA 3733

RESULT 5

US-10-764-390-6  
 ; Sequence 6, Application US/10764390  
 ; Publication No. US2004021421A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Agensys, Inc.  
 ; APPLICANT: Raitano, Arthur B.  
 ; APPLICANT: Jakobovits, Ava  
 ; APPLICANT: Challita-Bid, Pia M.  
 ; APPLICANT: Ge, Wangmao  
 ; APPLICANT: Faris, Mary  
 ; APPLICANT: Steven B. Kanner  
 ; APPLICANT: Juan J. Perez-Villar  
 ; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
 ; TITLE OF INVENTION: Entitled 254P1D6B Useful in Treatment and Detection of  
 ; TITLE OF INVENTION: Cancer  
 ; FILE REFERENCE: 51158-20081.00

; CURRENT APPLICATION NUMBER: US/10764,390  
 ; CURRENT FILING DATE: 2004-01-23  
 ; PRIOR APPLICATION NUMBER: US60/442,526  
 ; PRIOR FILING DATE: 2003-01-24  
 ; NUMBER OF SEQ ID NOS: 277  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 6  
 ; LENGTH: 6991  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (739)....(3930)  
 ; US-10-764-390-6  
 Alignment Scores:  
 Pred. No.: 0 Length: 6991  
 Score: 5527.00 Matches: 1068  
 Percent Similarity: 96.83% Conservative: 0  
 Best Local Similarity: 96.83% Mismatches: 4  
 Query Match: 99.05% Indels: 31  
 DB: 20 Gaps: 1  
 US-10-764-390-3 (1-1072) x US-10-764-390-6 (1-6991)  
 QY 1 MetalapProThrGlyValLeuSerSerLeuLeuLeuValThrIleAla----- 18  
 Db 621 ATGGCGCCCCCAGAGGTGCTCTCTTCATGTGCTGCTGGTGAATTTGACAGTTTGC 680  
 QY 18 ----- 18  
 Db 681 TTATGGTGGATGACTCATGCAAAAAAATCACTGGTGAGCATCATTTAAGAGAGCCCAT 740  
 QY 19 -----GlyCysAlaArgLysGlnCysSerGluGlyArgTh 30  
 Db 741 GACTAGACTGGCTGGCGAGCCCATGTTGTGCCGTAAAGCAGTGCAGCGAGGGGAGGAC 800  
 QY 30 rTyrSerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSerHi 50  
 Db 801 ATATTCCAATCGATGCTATTCACCTTAATCTGGAACCAACAGAAATCATGCGGGTGTCTCA 860  
 QY 50 sThrPheProValValAspCysThrAlaAlaCysCysAspLeuSerSerCysAspLeuAl 70  
 Db 861 CACCTTCCCTGTCTAGACTGCAGCGCGCTTGTGTGACCTGTCCAGCTGTGACCTGGC 920  
 QY 70 aTTPTrpPheGluGlyArgCysTyrLeuValSerCysProHisLysGluAsnCysGluPr 90  
 Db 921 CTGGTGGTTTCAGGGCGCTGCTACCTGCTGAGCTGCCCCCACAAGAGAACTGTGAGCC 980  
 QY 90 oLysLysMetGlyProLysArgSerTyrLeuThrPheValLeuArgProValGlnArgPr 110  
 Db 981 CAAGAAGATGGCGCCCATCAGTCTTATCTCACTTTTGTGCTCGCGCTGTTCAGAGGCC 1040  
 QY 110 oAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgGlySerProSerGlyIleTr 130  
 Db 1041 TGCACAGCTGCTGAGCTATGGGACATGATCTGAAACAGGGGCTCCCCCTCGGGATCTG 1100  
 QY 130 pGlyAspSerProGluAspIleArgLysAspLeu\*\*\*PheLeuGlyLysAspTrpGlyLe 150  
 Db 1101 GGGGACTCCTCCTGAGGATATCAGAAAGACTTGGCCCTTTCTAGGCAAGATTGGGGCT 1160  
 QY 150 uGluGluMetSerGluTyr\*\*\*AspAspTyrArgGluLeuGluLysAspLeuGluPr 170  
 Db 1161 AGAGGAGATGCTGAGTACTCAGATGACTACCGGAGCTGGAGAAGGACCTCTTGCAACC 1220  
 QY 170 oSerGlyLysGlnGluProArgGlySerAlaGluTyrThrAspTTPGlyLeuLeuProGl 190  
 Db 1221 CAGTGGCAAGCAGGAGGCCAGAGGAGTGGCCAGTACAGGACTGGGGCTTACTGCCGG 1280  
 QY 190 ySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThrGl 210  
 Db 1281 CAGCGAGGGGCTTCAACTCTCTGTTGGAGACAGTCTCTGCGGTGCCAGGAGACGCA 1340

QY 210 nGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLyLeuPr 230  
 Db 1341 GCAGGACCCTGAGCTCCATTACTGAATGAGTCGGCTTCAACCCCTGCCCCAAACTCCC 1400  
 QY 230 oGluArgSerValLeuLeuProLeuProThrThrProSerSerGlyGluValLeuGluY 250  
 Db 1401 TCAGAGAAAGTGTTGCTTCCCTTCCCGACTACTCCATCTTCAGGAGAGGTGTTCGAGAA 1460  
 QY 250 sGluLysAlaSerGlnLeuGlnGluGlnSerSerAsnSerSerGlyLysGluValLeuMe 270  
 Db 1461 AGAAAAGGGTTCTCAGCTCCAGAACATCCAGACACAGCTCTGAAAAGAGGTTCTAAT 1520  
 QY 270 tProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGluLysSe 290  
 Db 1521 GCCTTCCCATAGTCTTCCCTCCGCGCAAGCTCGAGCTCAGTCACTGCGAGAAAAG 1580  
 QY 290 rProValLeuThrValThrProGlySerThrGluHisSerIleProThrProProThrSe 310  
 Db 1581 CCCAGTGTCTCAGTCACTCCCGGGAGTACAGAGCACAGCATCCCAACACCTCCCACTAG 1640  
 QY 310 rAlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaProAr 330  
 Db 1641 CCAGCCCCCTCTGAGTCACCCCTCAGCTACCCATATCTCTACCACTGTCTCCAG 1700  
 QY 330 gThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAs 350  
 Db 1701 GACAGTGAAGAAGCTTACGGTATCGGCTCGAGATAACCTAAATATAACTTTTACCCGACAA 1760  
 QY 350 nGluValGluLeuLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTy 370  
 Db 1761 TGAAGTGAACCTGAAGGCCCTTGTTCGGCCAGCGCCACCTGTAGAAACAACCTACAACCTA 1820  
 QY 370 rGluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLy 390  
 Db 1821 TGAATGGAATTTAATAAGCCACCCACACACTACCAAGGTGAATAAATAAAGGACACAA 1880  
 QY 390 sGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSe 410  
 Db 1881 GCNAACCTTAACTCTCTCAATTGTCCTCGGACTTTATGCTTCAAGTCACTGTTTC 1940  
 QY 410 rSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgVa 430  
 Db 1941 TAGTGAAAACCGCTTTGGAGAAGGATTTGTCAATGTCACTGTAAGCCTGCCAGAAGAGT 2000  
 QY 430 lAsnLeuProValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuTh 450  
 Db 2001 CAACCTGCCACCTGTAGCAGTGTGTTTCTCCCAACTGCAAGAGCTCACTTGCCTTTGAC 2060  
 QY 450 rSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTr 470  
 Db 2061 GTCAGCCCTCATTTGATGGCAGCCAAAGTACAGNTGATACCTGAAATAGTGAGTTATCATTTG 2120  
 QY 470 pGluGluIleAsnGlyProPheIleGluGluLysThrSerValAspSerProValLeuAr 490  
 Db 2121 GGAAGAAATAAAGCGGCCCTTCATAGAAGAAGAAGCTTCAGTTGACTCTCCGCTCTTAGC 2180  
 QY 490 gLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGl 510  
 Db 2181 CTGTCTAACTTGTATCTCGGTAACTATATGTTTTCAGGTTGACTGTTTACAGACTCGGACCG 2240  
 QY 510 yAlaThrAsnSerThrThAlaAlaLeuIleValAsnAsnAlaValAspTyrProProVa 530  
 Db 2241 AGCCACTAACTCTACAACCTGCAGCCCTTAATAGTGAACAATGCTGTGCACTACCCACCT 2300  
 QY 530 lAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGl 550  
 Db 2301 TGTAAATGCAAGGACCAAAATCACACCACTTAATTTGCCCCCAAAACTCCATCACTTTGAATGG 2360  
 QY 550 yAnGlnSerSerAspAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySe 570  
 Db 2361 AAACCCAGAGCATGACATCACAGATTTGCTCTATGAGTGTCTCCCTGGGCTCTGGGAG 2420  
 QY 570 rGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMe 590

Db 2421 TGAGGGCAAAACATGTGTGTCATGCGGGAGTACAGACGCCATACCTTATTTTCTGCAAT 2480  
 QY 590 tGlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerTh 610  
 Db 2481 CGAGAAGGAGATATATACATTTTCAGCTGAAGGTGACAGATTTCTTCAAGGCAACAGTCTAC 2540  
 QY 610 rAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaGl 630  
 Db 2541 TGCTGTGTGACTGTGATTTGCCAGCTTGAACCAATAGACCTCCAGTGGCTGTGGCCGG 2600  
 QY 630 yProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerSe 650  
 Db 2601 CCCTGATAAAGAGCTGATCTTCCAGTGGAAAGTGTACCCCTGGATGGAGCAGCAGCAG 2660  
 QY 650 rAspAspHisGlyIleValPheTyrHisTrpGluHisValArgGlyProSerAlaValGl 670  
 Db 2661 CGATGACCACGGCATTTCTTCTACCACTGGGAGCAGCTCAGAGGCCCCAGTGCAGTGA 2720  
 QY 670 uMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHi 690  
 Db 2721 GATGGAAAATATTGCAAAAGCAATAGCCACTGTGACTGTCTCCAGGTGGGGACCTACCA 2780  
 QY 690 sPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAl 710  
 Db 2781 CTTCCGTTTGACAGTGAAGAACCCAGCGGCACTGAGCAGCAGCTCACCCCTCATCTGTGGC 2840  
 QY 710 aValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuVa 730  
 Db 2841 TGTGAAGAGGAAATATATAGTCTCTCCAGAGCCCGGCTGGTGGCAGACATGTTCTTGT 2900  
 QY 730 lLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleVa 750  
 Db 2901 GCTTCCCAATATTTCCATTACTTTTGGATGGTTCAAGGTCTACTGATGACCAAGAAATGT 2960  
 QY 750 lSerTyrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySe 770  
 Db 2961 GTCTCTCTGTGGATCCGGGATGGCCAGAGTCCAGCAGCTGGAGATGTCATCGATGGCTC 3020  
 QY 770 rAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLe 790  
 Db 3021 TGACACAGTGTGGCTCTGCAGCTTACGAATCTGTGTGAGGGGGTGTACATTTTCCACTT 3080  
 QY 790 uArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnPr 810  
 Db 3081 GCGAGTCAACCCAGCTCAGGGGGCTCGGACACAGACACTGCCACTGTGGAAGTCAGCC 3140  
 QY 810 oAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuTh 830  
 Db 3141 AGACCTTAGGAAGAGTGGCTGTGGAGCTGACCTGCAGGTTGGTGTGGGCAGCTGAC 3200  
 QY 830 rGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSe 850  
 Db 3201 AGACAGCGGAAGGACACCTTTGTGAGGACAGCTGGCTGTGCTGCTGAACAGTCTGGACTC 3260  
 QY 850 rAspIleLysValGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTy 870  
 Db 3261 GGACATTAAGGTCCAGAAGATTCGGGCCCACTCGGATCTCAGCACCGTATGTTGTTTA 3320  
 QY 870 rValGlnSerArgProProPheLysValLeuLysAlaAlaGluValAlaArgAsnLeuHi 890  
 Db 3321 TGTACAGCAGCGCCGCTTCAAGGTTCTCAAGCTGCTCAAGTGGGCCGGAATCTGCA 3380  
 QY 890 sMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspTh 910  
 Db 3381 CATCGGGCTCTCAAGGAGAGGCTGACTTCTTGTCTTTTCAAGGCTTGTAGGGTTGATAC 3440  
 QY 910 rAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCy 930  
 Db 3441 AGCAGGTTGCCTTCTGAAGTGTTCCTGGCCATGCTCACTGCGACCCCTCACAAAGGCTG 3500  
 QY 930 sIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGluSe 950

```

Db      3501 CATTTGCTCTCACTTATGGATGGAGAACCTTATACAGCGTTATATCTCGGATGGAGAG 3560
Qy      rAsnCysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuTh 970
|      |||||
Db      3561 CAACGTGTGAGTGGAGTATATCTATGTGACACAGTGTGGCTTTTACTCTTATTTGCTAAC 3620
Qy      rGlyGlyPheThrTrpLeuCysIleCysCysCysIleValGlnLysArgThrLysIleLea 990
|      |||||
Db      3621 AGGAGTTTCACTTGGCTTTGCACTGCTGCTGCAAAAGACAAAAGGACTAAATCAG 3680
Qy      gLysLysThrLysTyrThrIleLeuAspAsnMetAspGluGlnGluArgMetGluLeuAr 1010
|      |||||
Db      3681 GAAAAACAAAGTACACCATCTGGATAACATGGATGACACAGGAAAGAAATGGAACCTGAG 3740
Qy      gProLysTyrGlyLysIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerG1 1030
|      |||||
Db      3741 GCCCAAAATATGGTATCAAGACCCGAGCACAGACACAACTCCAGCCTGATGATCCGA 3800
Qy      uSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAs 1050
|      |||||
Db      3801 GTCTGAGTTTGACAGTGCACAGGACACAAATCTTCAGCCGAGAAAGATGGAGAGAGGAA 3860
Qy      nProLysValSerMetAsnGlySerIleArgAsnGlyAlaSerPheSerTyrCysSerLy 1070
|      |||||
Db      3861 TCCAAAGCTTTCATGATGTTCCATCAGAAATGGAGCTTCCTTCAGTTATTTGCTCAA 3920
Qy      1070 sAspArg 1072
|      |||||
Db      3921 GGACAGA 3927

RESULT 6
US-10-764-390-269
; Sequence 269, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Bid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Fatis, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 2541D6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10764,390
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 269
; LENGTH: 6991
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-764-390-269

Alignment Scores:
Pred. No.: 0 Length: 6991
Score: 5527.00 Matches: 1068
Percent Similarity: 96.83% Conservative: 0
Best Local Similarity: 96.83% Mismatches: 4
Query Match: 99.05% Indels: 31
DB: 20 Gaps: 1

US-10-764-390-3 (1-1072) x US-10-764-390-269 (1-6991)

Qy      1 MetAlaProThrGlyValLeuSerSerLeuLeuLeuValThrIleAla----- 18
|      |||||
Db      621 ATGGGCCCCCAGAGGTGCTCTCTTCATTGCTGCTGCTGGTGACAAATTGCAGTTTGC 680

```

```

Qy      18 ----- 18
Db      681 TTATGGTGGATGCACATCATGGCAAAAAAATCATCTGGTGAGCATCATTTAAGAAGACCCAT 740
Qy      19 -----GlyCyAlaArgLysGlnCysSerGluGlyArgTh 30
|      |||||
Db      741 GACTAGACTGGCTGGCGGAGCCCATGTTGTGCCGTAAGCAGTGCAGCGAGGGAGGAC 800
Qy      30 rTyrSerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSerHi 50
|      |||||
Db      801 ATATTCCAATGCAGTCAATTTCACTTAATCTGGAAAAACCAACAGAAATCATCGGGGTCTCTCA 860
Qy      50 sThrPheProValValAspCysThrAlaAlaCysCysAspLeuSerSerCysAspLeuAl 70
|      |||||
Db      861 CACCTTCCTCTGTCGAGACTGCACGGCGCTTGTGTGACCTGTCCAGCTGTGACCTGGC 920
Qy      70 aTrpTrpPheGluGlyArgCysTyrLeuValSerCysProHisLysGluAsnCysGluPr 90
|      |||||
Db      921 CTGTGTGTTTCAGGGCGCTGCTACCTACCTGGTGAGCTGCCCCCAAAAGAGAACTGTGAGCC 980
Qy      90 oLysLysMetGlyProIleArgSerTyrLeuThrPheValLeuArgProValGlnArgPr 110
|      |||||
Db      981 CAAGAAGATGGGCCCATCAGGTCTTATCTCCTTTGTGTCTCCGGCTGTTCAGAGGCC 1040
Qy      110 oAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgGlySerProSerGlyIleTr 130
|      |||||
Db      1041 TGCACAGCTGCTGACTATGGGGACATGATGCTGAACAGGGGGCTCCCTCCGGGATCTG 1100
Qy      130 pGlyAspSerProGluAspIleArgLysAspLeu***PheLeuGlyLysAspTrpGlyLe 150
|      |||||
Db      1101 GGGGACTCCTACCTGAGGATATCAGAAAGACTTGGCCCTTTCTAGGCAAGATTGGGGCT 1160
Qy      150 uGluGluMetSerGluTyr***AspAspTyrArgGluLeuGluLysAspLeuLeuGlnPr 170
|      |||||
Db      1161 AGAGGAGATGCTGAGTACTCAGATGACTACCGGAGCTGGAGAAGGACCTCTTGTCAACC 1220
Qy      170 oSerGlyLysGlnGluProArgGlySerAlaGluTyrThrAspTrpGlyLeuLeuProG1 190
|      |||||
Db      1221 CAGTGGCAAGCAGAGGCCAGAGGGAGTGGCGAGTACACGGACTGGGGCTTACTGCCGGG 1280
Qy      190 ySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThrG1 210
|      |||||
Db      1281 CAGCGAGGGGCCCTTCAACTCTCTGTTGGAGACAGTCTCGGGTGCAGCGGAGACCGCA 1340
Qy      210 nGlnAspProGluLeuHisTyrLeuAsnGlnSerAlaSerThrProAlaProLysLeuPr 230
|      |||||
Db      1341 GCAGGACCTCGAGCTCCATTACCTGAATGAGTGGCTTCAACCCCTGCCCAAACTCCC 1400
Qy      230 oGluArgSerValLeuLeuProLeuProThrThrProSerSerGlyGluValLeuGluY 250
|      |||||
Db      1401 TGAGAGAAAGTGTGTGCTTCCCTTTGCCGACTACTCCATCTTTCAGGAGAGGTGTTGGAGAA 1460
Qy      250 sGluLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLysGluValLeuMe 270
|      |||||
Db      1461 AGAAAAGCTTCTCAGCTCCAGGAAACATCAGCAACAGCTCTGGAAAGAGGTTCCTAAT 1520
Qy      270 tProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGluLysSe 290
|      |||||
Db      1521 GCCTTCCCATAGTCTTCTCCGGCAAGCTCGAGCTCAGCTCAGTCACCGTGGAGAAAAG 1580
Qy      290 rProValLeuThrValThrProGlySerThrGluHisSerIleProThrProThrSe 310
|      |||||
Db      1581 CCCAGTGTCCACAGTACCCCGGGGAGTACAGAGCAGACATCCCAACACCTCCACCTAG 1640
Qy      310 rAlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaProAr 330
|      |||||
Db      1641 GCGAGCCCCCTCTGAGTCCACCCCATCTGAGCTACCCATATCTCTACACTGCTCCAG 1700
Qy      330 gThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAs 350
|      |||||
Db      1701 GACAGTGAAGAACTTACGGTATCGGCTGGAGATAACCTAATATATACTTACCCGACAA 1760
Qy      350 nGluValGluLeuLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTy 370

```

1761 TGAAGTTGAACCTGAAGGCGCTTTGTTGCGCAGCGCCACCTGTAGAAAACAACCTACAACTA 1820  
 370 rGluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHis 390  
 1821 TGAATGGAAATTAATAAGCCACCCACAGACTACCAAGGTGAATATAAACCAAGGACACAA 1880  
 390 sGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSe 410  
 1881 GCAAACTCTTAACCTCTCTCAATTTGTCGTCCGACTTTATGTCTTCAAGTCACTGTTTC 1940  
 410 rSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgVal 430  
 1941 TAGTGAAACCGCTTTGGAGAAGGATTTGTCAATGTCACTGTAAGCCTGCCAGAAGAGT 2000  
 430 IAsnLeuProProValAlaValValSerProGlnLeuGlnLeuLeuThrLeuProLeuTh 450  
 2001 CAACCTGGCCACTGTAGCAGTTGTTTCTCCCACTGCAAGAGCTCACTTTGCCCTTTGAC 2060  
 450 rSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTr 470  
 2061 GTACGCCCTCATGTGAGCAGCCAAAGTACAGATGATGAAATAGTAGTATCATTTG 2120  
 470 pGluLeuIleAsnGlyProPheIleGluLysThrSerValAspSerProValLeuAr 490  
 2121 GGAAGAAATAAACGGGCCCTTCATAGAGAGAAGACTTCAGTTGACTCTCCCGTCTTACG 2180  
 490 gLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGl 510  
 2181 CTTGTCTAACTTGAATCCCTGGTAACATATAGTTTCAAGTTGACTGTGTACAGACTCGGACGG 2240  
 510 yAlaThrAsnSerThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProProVa 530  
 2241 AGCCACTAATCTACAACTGCGAGCCCTAATAGTGAACATGTGTGGACTACCCACAGT 2300  
 530 lAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGl 550  
 2301 TGCTAATGCGAGGACCAATACACCAATAACTTTGGCCCCAAAACCTCCATCACTTGAATGG 2360  
 550 yAsnGlnSerSerAspAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySe 570  
 2361 AAACCCAGACGATGACGATCACCAATATGCTCTATGATGTGCTCCCTGGGCTCTGGGAG 2420  
 570 rGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaWe 590  
 2421 TGAGGCAAAACATGTGGTTCATGCGAGGAGTACAGACCCATACCTTCATTTATCTGCAAT 2480  
 590 tGlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerTh 610  
 2481 GCAGGAAGGAGATTATACATTTCAAGGTGACAGATTTCTTCAAGGCCAACAGTCTAC 2540  
 610 rAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaGl 630  
 2541 TGCTGTGGTGACTGTGATTTGTCAGGCTGAAAACAAATAGACCTCCAGTGGCTGTGGCCGG 2600  
 630 yProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSe 650  
 2601 CCTGTATTAAGAGCTGATCTTCCAGTGGAAAGTGTCTACCTGGATGGAGGAGCAGCAG 2660  
 650 rAspAspHisGlyIleValPheTyrHisTrpGluHisValArgGlyProSerAlaValGl 670  
 2661 CGATGACCAACGGCATTTCTTACCACTGGGAGCAGCTCAGAGGCCCCAGTGCAGTGA 2720  
 670 uMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHi 690  
 2721 GATGGAAAAATATTGACAAAGCAATAGCCACTGTGACTGGTCTCCAGGTGGGGACCTACCA 2780  
 690 sPheArgLeuThrValIleAspGlnGlnGlyLeuSerThrSerThrLeuThrValAl 710  
 2781 CTTCCGTTTGACAGTGAAGACCAGCAGGAGCTGAGCAGCAGCTCCACCCCTCACTGTGGC 2840  
 710 aValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuVa 730

2841 TGTGAAGAAGAAATAATAAGTCTCCACAGAGCCCGGGCTGGTGGCAGACATGTTCTTGT 2900  
 730 lLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArqIleVa 750  
 2901 CTTTCCCAATATTTCCATTACTTTGGATGGTTCAAGGTCTACTGATGACCAAGATTTGT 2960  
 750 lSerTyrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySe 770  
 2961 GTCTCTATCTGTGGATCCGGATGGCCAGATCCAGAGTCCAGCAGTGGAGATGTCATCGATGGCTC 3020  
 770 rAspHisSerValAlaIleLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisIle 790  
 3021 TGACACACAGTGTGGCTCTGCAGCTTACGAAATCTGGTGGAGGGGGTGTACACTTTTCCACTT 3080  
 790 uArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnPr 810  
 3081 CGGAGTCCACACAGTCCAGGGGGCTCGGACACAGACACTGCCACTGTGGAGTGCAGCC 3140  
 810 oAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuTh 830  
 3141 AGACCTTAGGAAGAGTGGCTGGTGGAGCTGACCTGCAGGTGGTGGTGGCAGCTGAC 3200  
 830 rGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSe 850  
 3201 AGACGACGGAAGGACACCTTTGTGAGCAGCTGGCTGTGCTGGAACGCTGCTGGACTC 3260  
 850 rAspIleLysValGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyr 870  
 3261 GGACATTAAGGTCCAGAGATTTGGGCCCACTCGGATCTCAGCACCGTGTGTTGTTTA 3320  
 870 rValGlnSerArgProPropheLysValLeuLysAlaAlaGluValAlaAlaArgAsnLeuHi 890  
 3321 TGTACAGACGAGCGCGCTTTCAAGGTCTCAAAGCTGCTGAAGTGGCCCGAATACTGCA 3380  
 890 sMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspTh 910  
 3381 CATGGGCTCTCAAAGGAGAGGCTGACTTCTTGTCTTTTCAAGGTCTTTGAGGGTGTGATAC 3440  
 910 rAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCy 930  
 3441 AGCAGGTTGGCTTTTGAAGTGTCTGGCCATGTTCTGTCACCTGGTCACTCGCACCCCTCACAAGCGCTG 3500  
 930 sIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyLys 950  
 3501 CATTTGCTCTCATTTATGGATGGAGAACCTTATACAGCGTTATATCTGGGATGGAGAGAG 3560  
 950 rAsnCysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuTh 970  
 3561 CAACTGTGAGTGGAGTATATTTCTATGTGACAGTGTGGCTTTTACTCTTTATTTGTGCTAAC 3620  
 970 rGlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLysArgThrLysIleAr 990  
 3621 AGGAGGTTTCACTGGCTTTGCATCTGCTGTGCAAAAGACAAAAGAGACTTAAATCAG 3680  
 990 gLysLysThrLysTyrThrIleLeuAspAsnMetAspGluGlnGluArgMetGluLeuAr 1010  
 3681 GAAAAAACAAGATACACCTCTCGATAACATGATGAACAGGAAAGAAAGAACTGAACTGAG 3740  
 1010 gProLysTyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGl 1030  
 3741 GCCCAAAATATGGTATCAAGCACCGAAGCACAGACCAACTCCAGCCTGATGGTATCCGA 3800  
 1030 uSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyVa 1050  
 3801 GTCTGAGTTTGAAGTGAAGCAGGACCAATCTTCAGCCGAGAAAGATGAGAGAGGGA 3860  
 1050 nProLysValSerMetAsnGlySerIleArgAsnGlyAlaSerPheSerTyrCysSerby 1070  
 3861 TCCAAAGGTTTCCATGAATGGTTTCCATCAGAAATGGAGCTTCTTCAAGTATTGTCTCAAA 3920  
 1070 sAspArg 1072  
 3921 GGACAGA 3927

RESULT 7  
 US-11-097-143-16883  
 ; Sequence 16883, Application US/11097143  
 ; Publication No. US20050208558A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Venter, J. Craig  
 ; APPLICANT: et al  
 ; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
 ; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
 ; TITLE OF INVENTION: DROSOPHILA GENES.  
 ; FILE REFERENCE: CL000728  
 ; CURRENT APPLICATION NUMBER: US/11/097,143  
 ; CURRENT FILING DATE: 2005-04-04  
 ; PRIOR APPLICATION NUMBER: 60/157,832  
 ; PRIOR FILING DATE: 1999-10-05  
 ; PRIOR APPLICATION NUMBER: 60/160,191  
 ; PRIOR FILING DATE: 1999-10-19  
 ; PRIOR APPLICATION NUMBER: 60/161,932  
 ; PRIOR FILING DATE: 1999-10-28  
 ; PRIOR APPLICATION NUMBER: 60/164,769  
 ; PRIOR FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: 60/173,383  
 ; PRIOR FILING DATE: 1999-12-28  
 ; PRIOR APPLICATION NUMBER: 60/175,693  
 ; PRIOR FILING DATE: 2000-01-12  
 ; PRIOR APPLICATION NUMBER: 60/184,831  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: 60/191,637  
 ; PRIOR FILING DATE: 2000-03-23  
 ; NUMBER OF SEQ ID NOS: 43008  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16883  
 ; LENGTH: 3345  
 ; TYPE: DNA  
 ; ORGANISM: DROSOPHILA  
 US-11-097-143-16883  
  
 Alignment Scores:  
 Pred. No.: 1,08e-121 Length: 3345  
 Score: 1451.50 Matches: 373  
 Percent Similarity: 46.94% Conservative: 172  
 Best Local Similarity: 32.13% Mismatches: 383  
 Query Match: 26.01% Indels: 233  
 DB: 24 Gaps: 35  
  
 US-10-764-390-3 (1-1072) x US-11-097-143-16883 (1-3345)  
  
 Qy 8 LeuSerSerLeuLeuLeuValThrIleAlaGlyCysAla----- 21  
 Db 157 ATTTGCAACCTGCTTTTGTGCTGCCCACTCCGATGTCGTCAGCATATGCCGAGTCACAAAG 216  
 Qy 22 -----ArgLysGlnCysSerGluGlyArgThrTyr 31  
 Db 217 CAACACGCTGTTGGTGGAGTAAGAAACACAGGAGACCCCGCATTAATTCAGTT 276  
 Qy 32 SerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSerHisThr 51  
 Db 277 GCGGTTCT---ATATCGCCCAATTGTTGTGTCACCAAGATGCTGAGG-----CATGTG 327  
 Qy 52 Phe-----Pro 53  
 Db 328 TTCGAAAACGCCACTCCGCGGATGACGACGAGCGGGTGTGTTGAGGAGTACAAAGCCA 387  
 Qy 54 ValValAsp-----CysThrAlaAla 60  
 Db 388 CCGCGGATGATGAGGAGCCACTGGAGAGAGCGCTATCTTTGGAATGCTCTACAGCT 447  
 Qy 61 CysCysAsp-----LeuSerSerCysAspLeuAlaTrrpTrpPheGluGly 75  
 Db 448 TGCTGCGAAGAGCCCGGAATGGCAGCATGCTGCAATGTGTCTCGTCTTTTAAAGCC 507  
 Qy 76 ArgCysTyrLeuValSerCysProHisLysGluAsnCysGluProLys-----Lys 92

Db 508 AAATGCTACCATATCCGATGCCAGAGCAACGAGCCCTGTTGCCCAAGCTCAGAGTTCCG 567  
 Qy 93 MetGlyProIleArgSerTyrLeuThrPheValLeuArgProVal----- 107  
 Db 568 ATG---CCCAACGAAAGGTTTCAGATGGTACTGCTCAATCCACTGGCGGATGCCACATGG 624  
 Qy 108 -----GlnArgProAlaGlnLeuLeuLeuAspTyrGlyAsp 118  
 Db 625 CCACAGCTCTCAAGCGGAGGACCAACAGAAATGCTGAGATTCTTCCATACGATGAA 684  
 Qy 119 MetMetLeuAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArg 138  
 Db 685 GCGGCACCTTAAT-----TTCGGAAGACGAGCCC----- 711  
 Qy 139 LysAspLeu\*\*\*PheLeuGlyLysAspTrpGlyLeuGluGluMetSerGluTyr\*\*\*Asp 158  
 Db 712 AGAAGGTTGAGCTACTGCTCGCAAT-----CAGGAACCCCGTTTATGAGGAC 762  
 Qy 159 Asp-----TyrArgGluLeuGlyLysAspLeuLeuGlnProSerGlyLys 173  
 Db 763 GAAGATTTCCTTGGCCGACAGCAATGAATCAATGATCTTCCAAACCCGACGAAAC 822  
 Qy 174 GlnGluProArgGlySerAlaGluTyrThrAspTrpGlyLeuLeuProGlySerGluGly 193  
 Db 822 ----- 822  
 Qy 194 AlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThrGlnGlnAspPro 213  
 Db 823 -----GATGTTCTGCCCAACGAA 840  
 Qy 214 GluLeuHisTyrLeuAsnGluSerAla-----SerThrProAla 226  
 Db 841 GAGCTCGGCTACTACGATTCGAATGCCAAGTTTACCACCTCGCATATGGAGACACCTTGT 900  
 Qy 227 ProLysLeuProGluArgSerValLeuLeuProLeuProThrThrProSerSer----- 244  
 Db 901 CCACCA---CCGCAACAAATGTGTA-----CCACTACAGCCAAATGCCGTGCGT 945  
 Qy 245 -----GlyGluValLeuGluLysGluLys----- 252  
 Db 946 GGAGTGTGCACCTGTCCAGAGGCTTTGTGGGAATAAACAAGAGAGTGGTAATGGCA 1005  
 Qy 253 -----AlaSerGlnLeuGlnGluSerSerAsnSerSerGlyLysGluVal 268  
 Db 1006 GCGGTTCCCTATAGCTCATATCTA-----ACCAGCAATGAGCGGACAGCAAGAA 1056  
 Qy 269 LeuMetProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGlu 288  
 Db 1057 GCGGCTGCTAGTGAGAACTCACCGAGGTATCA----- 1089  
 Qy 289 LysSerProValLeuThrValThrProGlySerThrGluHisSerIleProThrProPro 308  
 Db 1090 -----ACGCGGCT 1098  
 Qy 309 ThrSerAlaAlaProSerGluSerThrThrProSerGluLeuProIleSerProThrThra 328  
 Db 1099 TTGAAGGCTGAGCAGAAC----- 1116  
 Qy 329 ProArgThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuPro 348  
 Db 1117 -----AAGGACATTGTAGTCTCCGTAATGTCCAAG---GAAGTTGTTGCTGCG 1161  
 Qy 349 AspAsnGluValGluLeuLysAlaPheValAlaProAlaProVal---GluThrThr 367  
 Db 1162 GAACAGAGGTGACCTTAGCGGCTTTACGTTCCCGATGAGCAACACGAGGATACCAAG 1221  
 Qy 368 TyrAsnTyrGluTrpAsnLeuIleSerHisProThr---AspTyrGlnGlyGluIleLys 386  
 Db 1222 TACAAATATCTTGGACACTTATATCGCAACCCCAAGGGTCTCTATGAATGGCACTATTTCG 1281  
 Qy 387 GlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLys 406

Db 1282 GATCAGACGAAGTCGAAGTAAACTATCGAATCTTTTCGGAGGAGCTGTATACTTTTAAAG 1341  
Qy ValThrValSerSerGluAsn---AlaPheGlyGluGlyPheValAsnValThrValLys 425  
Db 1342 GTCACTGTAACTGTGTGACAATGGAACTTTGGCGAGGCAACGCAATGTTACAGTGT 1401  
Qy 426 ProAlaArgValAsnLeuProProValAlaValValSerProGlnLeuGlnGluLeu 445  
Db 1402 CCCGAGAATCGAATCAATCAGCCACCAAGTCATCATCTGCCCCAGGAGCAGATCATC 1461  
Qy 446 ThrLeuProLeuThrSerAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 465  
Db 1462 CGTCAGCCACCAACCAATGCTATCTGATGGCAGCAGCAGTACGATGATGATGAATC 1521  
Qy 466 ValSerTyrHisTrpGluGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 485  
Db 1522 ACCAATTTGGCTGGAGGTGATTTCCGGACCGATGGTTACCAACCAAGTTTTCGCAGAG 1581  
Qy 486 SerProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrVal 505  
Db 1582 GTTAACACCCCTCAGTGGATCTCATATCGCCGGGGAACACATTCAGAGTAAACGTA 1641  
Qy 506 ThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuLeuValAsnAsnAlaVal 525  
Db 1642 ACCGATTCAAATAACGTGACCAATTCACCACTGCCAGCATAGCAGTCTTTAAAGGAGACC 1701  
Qy 526 AspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuLeuProGlnAsnSer 545  
Db 1702 GATTATGCTCCGGTGGCAATCCCGCGAGTCCGTGATCTTGATTTGCCCAACCAATAAT 1761  
Qy 546 IleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuLeuTrpSer 565  
Db 1762 GTCACTCTGAATGGCAGACCCAGTTCGGATGATCAGAGATGTTGCTTGGAGTGGACC 1821  
Qy 566 LeuGlyProGlySerGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeu 585  
Db 1822 AAGGATGCTAGCGACGACCAAGCTGTGGATATGCAATATACAGAAACACCTATGTT 1881  
Qy 586 HisLeuSerAlaMetGlnGlyAspTyrThrPheGlnLeuLysValThrAspSerSer 605  
Db 1882 CAGCTGTCCAAATTTGGAGAGGCGCATGTACACTTTTGTCTAAAGTAACCGATGGCAGT 1941  
Qy 606 ArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProPro 625  
Db 1942 GGGCAATCGAGTACAGTAAAGTCCATGTGTTGTGAAGCTCCCAAGATTTCTCCACCA 2001  
Qy 626 ValAlaValAlaGlyProAspLysGluLeuIlePheProValGluSerAlaThrLeuAsp 645  
Db 2002 GTTGCTGAGCGAGGAAGTAATACGACTACTAGCTTGCCCATTAATTTGGGTTCTTTTGAAT 2061  
Qy 646 GlySerSerSerSerAspAspHisGlyIleValPheTyrHisTrpGluHisValArgGly 665  
Db 2062 GGCTCCGATTCGAAGACACATATGGCATCAAGAGTTATTTGTGAAGCAGCTGAGCGGG 2121  
Qy 666 ProSerAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGln 685  
Db 2122 CCCAATAATCGAGTCATTTTGAAGTCCAATCATCAATGCGCAATCCACTCTCTGACT 2181  
Qy 686 ValGlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSer 705  
Db 2182 CTGGGACTCTATGAGTTCCGAATTAACCGTAGCTGATGAAAAATAATAACACCGCTACGGAC 2241  
Qy 706 ThrLeuThrValAlaValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyGly 725  
Db 2242 ACCAGTGGTGAAGATAGTTCAAGAACCGAACCGACGCTCCCAATAGCCAAACCGCGTGT 2301  
Qy 726 ArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAsp 745  
Db 2302 GATCACACCGTCACTCCCTGGCGCCACCGCATCTATTTCAATGGCTCTAAATCCTGGAT 2361  
Qy 746 AspGlnArgIleValSerTyrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAsp 765  
Db 2362 GACCTGGCTGTGTCAAGTATCTCTGGACACGCGAGCAGCAGATTTGGCAGCGGCGCTC 2421

Qy 766 ValIleAspGlySerAspHisSerValAlaLeuLeuGlnLeuThrAsnLeuValGluGlyVal 785  
Db 2422 ATTGTGGCAGATACCGACAGGAGCCCGTAATGATTTTGACCAATTTAGTACAGCGCGC 2481  
Qy 786 TyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThr 805  
Db 2482 TATGTATTCACACTGACTGTGAGCGATGATCAGGGTTTGACCAAGTTCAGATACTGTTCAG 2541  
Qy 806 ValGluValGlnProAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGly 825  
Db 2542 GTTAATGTCGTCGCGATCCCAAGCTATGTAATTTGGTTCAATGACCCCTGCCCATGGC 2601  
Qy 826 ValGlyGlnLeuThrGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeu 845  
Db 2602 ATCTCTGTCTCGTCAATCCGAACCTGGATTCGGTGGTGCAAAACTGCAGTTGTTGCTG 2661  
Qy 846 AsnValLeuAspSerAspIleLysValGlnLysIleArg-----AlaHisSerAspLeu 863  
Db 2662 GGAGAT---GAGAATAAGATCCAGTCCAGGAGCTAAAAATATGACTTCATACCGAT--- 2715  
Qy 864 SerThrValIleValPheTyrValGlnSerArgProProPheLysValLeuLysAlaAla 883  
Db 2716 GCTACTGTCCTGGTCTTCTACGTAAACGATGACAGGGT---AAGGCAATGGATGGTTG 2772  
Qy 884 GluValAlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAspPheLeuLeuPhe 903  
Db 2773 CAAGTGGAGCGCAGTTGAGAACCACTCAGAAAGGATGATCTATACTGGCGCATTT 2832  
Qy 904 LysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCys 923  
Db 2833 GCAGTG---GATAATCGCACCTCTGTTTGCAGAGTGAATGCTCCGGTCATGGGTCGT 2889  
Qy 924 AspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArg 943  
Db 2890 AATCCCAATTACAGAGCTTGATATGCGAGCGTTCGGATGCCATCG-----GCTGGC 2943  
Qy 944 TyrlleTrpAspGly---GluSerAsnCysGluTrpSerIlePheTyrValThrValLeu 962  
Db 2944 TACTTCTTATTAACAGGAGCGCCAAATGCGACTGCTCCATATATATATATGTTGTTGTT 3003  
Qy 963 AlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLys 982  
Db 3004 GTGATTGGGCTGCTCTCTGCTATCTGAGTATTTTGGGCGATAGCCTGCTGCTGCAGA 3063  
Qy 983 ArgGlnLysArgThrLysIleArgLysLysThr---LysTyrThrIleLeuAspAsnMet 1001  
Db 3064 CAATCGAAAAAGCCCGCTTCTCGTCAAGAAAGTGCAAAAATATTCATTGATTGTAATAAG 3123  
Qy 1002 AspGluGlnGluArgMetGluLeuArgProLysTyrGlyIleLysHisArgSerThrGlu 1021  
Db 3124 GATGAAGAGCAGCC----- 3138  
Qy 1022 HisAsnSerSerLeuMetValSerGluSerGluPheAspSerAspGlnAspThrIlePhe 1041  
Db 3139 ---AATTATTCTCGAAACACATCGCTGACGGAATCAGAGCGAATTCGATGCTCTTTT 3195  
Qy 1042 SerArgGluLysMetGluArgGlyAsnProLysValSerMetAsnGlySerIleArgAsn 1061  
Db 3196 GAGACCCGACCAAGTCCAAATGGTCTGGGCAAGCAAGTCCGACCAATTTCTCACAGTCA 3255  
Qy 1062 Gly 1062  
Db 3256 GGA 3258

## RESULT 8

US-10-450-763-8985  
; Sequence 8985, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIF3/US



; CURRENT APPLICATION NUMBER: US/10/450,763  
 ; CURRENT FILING DATE: 2003-06-11  
 ; PRIOR APPLICATION NUMBER: PCT/US01/08631  
 ; PRIOR FILING DATE: 2001-03-30  
 ; PRIOR APPLICATION NUMBER: 09/540,217  
 ; PRIOR FILING DATE: 2000-03-31  
 ; PRIOR APPLICATION NUMBER: 09/649,167  
 ; PRIOR FILING DATE: 2000-08-23  
 ; NUMBER OF SEQ ID NOS: 60736  
 ; SOFTWARE: Custom  
 ; SEQ ID NO 8985  
 ; LENGTH: 2423  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SIMILAR  
 ; LOCATION: (479)..(1723)  
 ; OTHER INFORMATION: 60% homologous to Homo sapiens dJ73M23.3 (KIAA0319), accession  
 ; OTHER INFORMATION: number AL031230, Smith-Waterman Score=1025.  
 US-10-450-763-8985

Alignment Scores:  
 Pred. No.: 2,52e-109 Length: 2423  
 Score: 1314.00 Matches: 358  
 Percent Similarity: 53.20% Conservative: 41  
 Best Local Similarity: 47.73% Mismatches: 164  
 Query Match: 23.55% Indels: 188  
 DB: 22 Gaps: 23

US-10-764-390-3 (1-1072) x US-10-450-763-8985 (1-2423)

Qy 365 GluThrThrTyrAsnTyrGluThrAsnLeuLeuSerHisProThrAspTyrGlnGlyGlu 384  
 Db |||||  
 Qy 62 GAAACAACCTACAACTATGAATGGAATTTAATAGCCTCTACGTAGTGGACCCATCAGT 121  
 Db |||||  
 Qy 385 IleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrVal 404  
 Db |||||  
 Qy 122 CTTGCATGTTGAAGTACAACTTTCTTTATGTTTACAGTTGTCCGTGGACTTTATGTC 181  
 Db |||||  
 Qy 405 PheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrVal 424  
 Db |||||  
 Qy 182 TTCATGTG-----GGTAAGACATATGCCCTACACAGATGGGCTCAGTCACTGTC 232  
 Db |||||  
 Qy 425 -----LysProAlaArgValAsnLeuProProValAlaValValSerProGln 441  
 Db |||||  
 Qy 233 CTTTTTTTGTGTACGCCAGAGAGTCAACCTGCCACCTGTAGCAGTTTATAAACTTATG 292  
 Db |||||  
 Qy 442 LeuGlnGluLeuThrLeuProLeuThrSerAlaLeuLeuLeuAspGlySerGlnSerThrAsp 461  
 Db |||||  
 Qy 293 TCTTCCTCTCTGTGATTCTTCTTGTCTTTTATTTCTTAATCTTTGTAGAAAGTACAGAT 352  
 Db |||||  
 Qy 462 AspThrGluIleValSerTyrHisTyrGluGluLeuLeuAsnGlyProPheIleGluGluLys 481  
 Db |||||  
 Qy 353 GATACGTAATAGTAGGTATCAITGGGAAGAAATAAACGGGCCCTTCATAGAGAGAG 412  
 Db |||||  
 Qy 482 ThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPhe 501  
 Db |||||  
 Qy 413 ACTTACTTGTGCTC-----TTTACTTTT 436  
 Db |||||  
 Qy 502 ArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaLeuLeuVal 521  
 Db |||||  
 Qy 437 TTGTTTTGTTTT----- 448  
 Db |||||  
 Qy 522 AsnAsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeu 541  
 Db |||||  
 Qy 448 ----- 448  
 Qy 542 ProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeu 561  
 Db |||||  
 Qy 449 -----GTTTTTATT 457  
 Db |||||  
 Qy 562 TyrGluTrpSerLeuGlyProGlySerGluGlyTyrHisValValMetGlnGlyValGln 581  
 Db |||||

Db 458 TTTGTATTGCT-----ATTACTTAGGAGTACAG 487  
 Qy 582 ThrProTyrIleuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLysVal 601  
 Db |||||  
 Db 488 ACGCATACCTTCTATTCTGCAATCGAGAGAGAGATTATACATTTTCAGCTCAGGGA 547  
 Qy 602 ThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnProGluAsn 621  
 Db |||||  
 Db 548 TTGCAGACTTAATGCAAGAGAGATAATCTTTTTCACACTGTTTCTGTAAATTCAGAAAAC 607  
 Qy 622 AsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGluSer 641  
 Db |||||  
 Db 608 AATAGACTCCAGTGGCTGTGGCGCCCTGATAAAGAGCTGATCTTCCAGTGGAAAGT 667  
 Qy 642 AlaThrLeuAspGlySerSerSerSerAspAspHisGly----- 654  
 Db |||||  
 Db 668 GCTACCTGTC-----CTGTCAATATTCAAAACCCCATGCCCTCCCTTATGTAAACCACT 721  
 Qy 655 IleValPheTyrHisTyrGluHisValArgGlyProSerAlaValGluMetGluAsnIle 674  
 Db |||||  
 Db 722 TCTATGTTTTAT-----AGAGGCCCCAGTGCAGTGGAGATGGAATAATATT 766  
 Qy 675 AspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThr 694  
 Db |||||  
 Db 767 GACAAGCAATAGCACTGTGACTGGTCTCCAGGTGGGACCTACCCTTCGTGTTT--- 823  
 Qy 695 ValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaVal----- 711  
 Db |||||  
 Db 824 -----TATTCTATCACTGGTGTGTTTAAATCAATCTCTTGTCTTTCAT 865  
 Qy 712 -----LysLysGluAsnAsnSerProProArgAlaAArgAlaGlyArgHis 727  
 Db |||||  
 Db 866 TTTGCTATATTGACTAGTAAATAATAGTCTCCAGAGCCGGGCTGGGGCAGACAT 925  
 Qy 728 ValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAsp----- 745  
 Db |||||  
 Db 926 GTTCTTGTGCTTCCCAATAATTCATCTTTGATGGTTCAGGTCTAAGAGGCCAG 985  
 Qy 746 -----AspGlnArgIleValSerTyrLeuTrpIleArgAspGlyGlnSerProAlaAla 763  
 Db |||||  
 Db 986 AATTGGGACCCGAGATTGACAGCTGCCCTGTGATCCGT-----GTTCTTTTCG 1033  
 Qy 764 GlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGlu 783  
 Db |||||  
 Db 1034 CAGGATGTCATCGATGGCTCTGACACAGTGGCTCTGCAGCTTACGAAATCTGGTGAG 1093  
 Qy 784 GlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAspThr 803  
 Db |||||  
 Db 1094 GGGGTGTACACTTTCCACTTCTTATGCTCTAAAGCCTGTGCAAAATGTCGAAGTCAAAAT 1153  
 Qy 804 AlaThrValGluValGlnProAspProArgLysSerGlyLeuValGluLeuThrLeuGln 823  
 Db |||||  
 Db 1154 CTTGTATCTTCTTTTCCCGACAGCCCTAGGAAGAGTGGCTGTGGAGCTGACCTGCGAG 1213  
 Qy 824 ValGlyValGlyGlnLeuThrGluGlnArgLysAspThrLeuValArgGlnLeuAlaVal 843  
 Db |||||  
 Db 1214 GTTGGTGTGGGAGCTGACAGAGCAGCGGAAGACACCTTGTGTAGGAGCTGGCTGTG 1273  
 Qy 844 LeuLeuAsnValLeuAspSerAspIleValGlnLysIleArgAlaHisSerAspLeu 863  
 Db |||||  
 Db 1274 GCGGTGTATGGTATGTCACCATCTGCACCATGTG---GCTGCTCTCTCATTTGCTCTC 1330  
 Qy 864 ---SerThrValIleValPheTyrValGlnSerArgProProPheLysValLeuLysAla 882  
 Db |||||  
 Db 1331 CACAGACCGGATGTTGTTTTATGTACAGAGAGCGCCCTTTCAGAGTTCTCAAGCT 1390  
 Qy 883 AlaGluValAlaArgAsnLeuHisMet--ArgLeuSerLysGlyLysAlaAspPheLeu 902  
 Db |||||  
 Db 1391 GCTGAAGTGGCCGAAATCTGCACATAGCTGAATCCAGGGATTCCTCTTCTGCTGCTGCT 1450  
 Qy 902 uPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHis 922  
 Db |||||  
 Db 1451 GCCCTCAGCCCTACACACCTCTTT--GCAGGTTGCTTCTGAAGTGTCTGAGCCATGGTCA 1509

```

QY 922 sCysAspProLeuThrLysArgCysile-----CysSer----- 933
Db 1510 CTGCGACCCCTCACAAAGCGCTGATTGAAATAAAGTTGTTCTACAAATAGCCTGTA 1569
QY 934 -----HisLeuTrpMetGluAsnLeuLeuGlnArgTyrIleTrpAspGlyGluSerAs 951
Db 1570 ACTATGTCACACTTGGTTT-----TC 1590

QY 951 nCys-----GluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValle 969
Db 1591 TTGTTTTCAGAGTGGAGTATATTCTATGTGACGTGTGAATCCTAGCTCTCAGTT-- 1648
QY 969 uThrGlyGlyPheThrTrpLeuCysile-----CysCysLysLysArgG1 984
Db 1649 -----ACTTTCTAAAGATCTGAGTTGAGCTTTTCATGTTCCTTTTCAGACA 1695

QY 984 nLysArgThrLysIleArgLysLysThr-----LysTyr-- 995
Db 1696 AAAAAGGACTAAATCAGAAAAAACGTTGAATATGTACTAGGTACATCTCAAGTACTT 1755
QY 996 -----ThrIleLeuAsp----- 999

Db 1756 TACCATTGAATCTAATCTCTCCCATAGCGGTACGAGTATGTGCAATGATTATCCCA 1815
QY 1000 -----AsnMetAspGluGlnGluArgMetGluLeuArgProLysTy 1013
Db 1816 CGACTGACAGCTGGAAGTGAATCAGCGGTTCTCTCTCAGATGCTTCTCTCTCT 1875

QY 1013 rGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGluSerGluPh 1033
Db 1876 AGGTATCAAGCACCGAAGCACAGACACAACTCCAGCTGATGTATCCGAGTCTGTGAG 1935

QY 1033 eAspSerAspGlnAspThrIlePheSer 1042
Db 1936 CAGTTTTCAGAGGAGTGTCTGCAAAAGC 1963

```

# RESULT 9

```

US-11-097-143-16882
; Sequence 16882, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16882
; LENGTH: 5775
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-16882

```

```

Alignment Scores:
Pred. No.: 1.05e-105 Length: 5775
Score: 1280.50 Matches: 384
Percent Similarity: 42.89% Conservatism: 171
Best Local Similarity: 29.68% Mismatches: 385
Query Match: 22.95% Indels: 359
DB: 24 Gaps: 43

US-10-764-390-3 (1-1072) x US-11-097-143-16882 (1-5775)

QY 8 LeuSerSerLeuLeuLeuValThrIleAlaGlyCysAla----- 21
Db 1157 ATTTGCCAACCTCTTTTGTGTCACATGCGATGTGTCAGCATATGCGACGTCACAAG 1216
QY 22 -----ArgLysGlnCysSerGluGlyArgThrTyr 31
Db 1217 CAAAAGCGCTGTGGTTGGAAGTAAGAAACACAAAGAGAGACCGCCCGCATATATTCAGTT 1276
QY 32 SerAsnAlaValIleSerProAsnLeuGluThrArgIleMetArgValSerHisThr 51
Db 1277 GCGGTTCT--ATATCGCCCAATTTGCTGCCACAAAGATGCTGAGG-----CATGTG 1327
QY 52 Phe-----Pro 53
Db 1328 TTGAAAACGCCCACTCCGCGGATGAGCAGCAGCGGGTGTGTTGAGGAGTACAAGCCA 1387
QY 54 ValValAsp-----CysThrAlaAla 60
Db 1388 CCGCGGATGAGTGGAGCCACTGGAGGAGGAGGCGCTATCTTTGGAAGTCCCTACAGGT 1447
QY 61 CysCysAsp-----LeuSerSerCysAspLeuAlaTrpTrpPheGluGly 75
Db 1448 TGCTGGAAAAGCCCCGGAATGGCAGCAGTGCCTGCAATGTGGTCTCGTTTAAAGGCC 1507
QY 76 ArgCysTyrLeuValSerCysProHisLysGluAsnCysGluProLys-----Lys 92
Db 1508 AAATGCTACCATATCCGATGCCAGAGCAACGAGGCGCTGTTGCCCAAGCTCAGAGTTCG 1567
QY 93 MetGlyProIleArgSerTyrLeuThrPheValLeuArgProVal----- 107
Db 1568 ATG---CCCAACGAAAGGTTTCAGATGCTATGTTGTCATCCATCCAGTGGGCGATGCCACATGG 1624
QY 108 -----GlnArgProAlaGlnLeuLeuAspTyrGlyAsp 118
Db 1625 CCACAGCTCTCAGGCGGAGGAGCCAAACAGATGCTGAGATTTCTGCCATACCATGAA 1684
QY 119 MetMetLeuAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArg 138
Db 1685 GCGGCACCTAAT-----TTCTGGAAGCAGGCC----- 1711
QY 139 LysAspLeu***PheLeuGlyLysAsp-----TrpGlyLeuGlu----- 151
Db 1712 AGAAGGTTGAGTACCTCGCTCGCAATCAGGTAATAGCATATGTTGTTCTCTCTATATTC 1771
QY 152 -----GluMetSerGluTyr***AspAsp 159
Db 1772 CTGACTATTATTAATCTTATAACATTTCTTGTAGGAAACCCCGTTTATGAGGACGAA 1831
QY 160 -----TyrArgGluLeuGlyLysAspLeuLeuGlnProSerGlyLysGln 174
Db 1832 GATTTCCTTGGCGACAGCAAGGAATGAATCAATGATCTTCCACCCGACGAAAAAC--- 1888
QY 175 GluProArgGlySerAlaGluTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAla 194
Db 1888 ----- 1888
QY 195 PheAsnSerSerValGlyAspSerProAlaValProAlaGluThrGlnGlnAspProGlu 214
Db 1889 -----GATGTTCTGCGCCAAACGAGAG 1909
QY 215 LeuHisTyrLeuAsnGluSerAla-----SerThrProAlaPro 227

```

Db 1910 CTCGGCTACTACGATTGCAATGCAAGTTTACCACTGCGGATATGGAGACACCTTGTCCA 1969  
 Qy 228 LysLeuProGluArgSerValLeuLeuProLeuProThrProThrProSerSer----- 244  
 Db 1970 CCA---CGCAACAATGTGA-----CCACTACAGCCAAATGCGGTGCGTGA 2014  
 Qy 245 -----GlyGluValLeuGluLysGluLys----- 252  
 Db 2015 GTGTGCACCTGTCAGAAAGGCTTTGTGTGGAAATAAACAAGGAAGTGCCTAATGGCAGCG 2074  
 Qy 253 -----AlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLysGluValLeu 269  
 Db 2075 GTTCCCTATAGCTCATATCTA-----ACCAGCAATGAGCGGGCAGACCAAGAAGCG 2125  
 Qy 270 MetProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGluLys 289  
 Db 2126 GCTGCTAGTGAGAACTCACCGAGGTATCA----- 2155  
 Qy 290 SerProValLeuThrValThrProGlySerThrGluHisSerIleProThrProThr 309  
 Db 2156 -----GlyGluValLeuThrValThrProThrProThrProThrProThr 319  
 Qy 310 SerAlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrAlaPro 329  
 Db 2168 AAGCTCAGCAGAAC----- 2182  
 Qy 330 ArgThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleLeuLeuProAsp 349  
 Db 2183 -----AAGGACATTTAGTCTCGTAAATGTCGAAG---GAAGTTGCTTTGCCGGA 2230  
 Qy 350 AsnGluValGluLeuLysAlaPheValAlaProAlaProProVal---GluThrThrTyr 368  
 Db 2231 CAGGAGGTGACCTTAGCGGGCTTTACGGTTCGATGCGCAACACCGCATACCAAGTAC 2290  
 Qy 369 AsnTyrGluTrpAsnLeuIleSerHisProThr---AspTyrGlnGlyGluLeuLysGln 387  
 Db 2291 AAATATCTTTGGACACTTATATCCCAACCCCAAGGCTCTATGAATGCGACTTTTCGGAT 2350  
 Qy 388 GlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysVal 407  
 Db 2351 CAGAGCAAGTCGAAGGTAAATATCGAATCTTTCCGGAGGAGCTGTATCTTTTAAGGTC 2410  
 Qy 408 ThrValSerSerGluAsn---AlaPheGlyGluGlyPheValAsnValThrValLysPro 426  
 Db 2411 ACTGTAACTGTGCATATGGAACCTTTGGCAGGCAACAGCAATGTTTACAGTCTCTCCC 2470  
 Qy 427 AlaArgValAlaLeuProProValAlaValValSerProGlnLeuGlnGluLeuThr 446  
 Db 2471 GAGAAATCGAATCAATCAGCCACCAAGTCATCTCGCCAGGGAGCATCATCCGT 2530  
 Qy 447 LeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleVal 466  
 Db 2531 CAGCCCAACCAATGCCATATCTGATGGCAGCAGCAGTACGATGATGAATCACC 2590  
 Qy 467 SerTyrHisTrpGluGluLeuAsnGlyProPheIleGluLysThrSerValAspSer 486  
 Db 2591 AATTGGCACTGGGAGGTGATTTCCGGACCGATTGGTTTACCAACAGTTTTCGACAGGTT 2650  
 Qy 487 ProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThr 506  
 Db 2651 AACACCTTCCAGTTGGATCTCACAATCGCCGGGAACTACACATTCAAGCTTAACCGTAACC 2710  
 Qy 507 AspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuIleValAsnAsnAlaValAsp 526  
 Db 2711 GATTCAATAACGTGACCAATTCACCACTGCCAGCATAGCAGTTCTTAAGGAGACCGAT 2770  
 Qy 527 TyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIle 546  
 Db 2771 TATGTCGGTGGCAATCGCGGGATGCGCGGATCTGTGATCTTGTATTTGCCCAACAATAATGTC 2830  
 Qy 547 ThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuTyrGluTrpSerLeu 566  
 Db 2831 ACTCTGAATGGCACAGCCAGTTCCGATGATCAGAGATCGTTGCTTGGGAGTGACCAAG 2890

Qy 567 GlyProGlySerGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHis 586  
 Db 2891 GATGCTAGCCAGCAGGCGCAAGGCTGTGATATGAGAAATACAAGAACACCTATGTTGAG 2950  
 Qy 587 LeuSerAlaMetGlnGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArg 606  
 Db 2951 CTGTCCAAATTTGGAGGAGGCGATGTACACATTTTGTCTAAAGTAAACGATGGCAGTGGG 3010  
 Qy 607 GlnGlnSerThrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProVal 626  
 Db 3011 CAATCGAGTACAGTAAGGTCCATGTGTTGTGAAGCTCCACGAATTTCTCCACCAAGTT 3070  
 Qy 627 AlaValAlaGlyProAsp----- 632  
 Db 3071 GCTGAGCAGGAAGTAAATACGGTAGGTTTCAAGTAAAGATATATTTATTCGTTTAA 3130  
 Qy 633 -----LysGluLeuPheProValGluSerAlaThrLeuAspGlySer 647  
 Db 3131 TATGATCTGTTCACAGACTACTAGC-TTGCCTAATTAATTTGGGTTCTTTGAATGGTCC 3189  
 Qy 648 SerSerSerAspAspHisGlyIleValPheTyrHisTrpGluHisValArgGlyPro--- 666  
 Db 3190 GATTGGAAGGACGACATTTGGCATCAAGAGTTATTTGTGAAGCAGCTGAGCGGCGCCAAT 3249  
 Qy 667 SerAlaValGluMetGlu----- 672  
 Db 3250 AATGCTAGCTATTTGAAGTCCAACTCATCAGTAAGTCATATATATATATATATATGTTG 3309  
 Qy 673 -----AsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGly 687  
 Db 3310 GACCTCATGCACTCCCATTTAACCTGCAGATTGCCAATGCCACCTCTCTGACTCTGGA 3369  
 Qy 688 ThrTyrHisPheArgLeuThrValLysAspGlnGlyLeuSerSerThrSerThrLeu 707  
 Db 3370 CTCTATGAGTTCGAATTAACCGTAGCTGATGAAATAATAACACCGCTACGCACCAACG 3429  
 Qy 708 ThrValAlaValLysLys----- 713  
 Db 3430 TGGGTGAAGATAGT-TCAAGGTAATCATATCCATAATCTTTCAGTAGTCTATAGTGCTAA 3488  
 Qy 714 -----GluAsnAspSerProProArgAlaArgAlaGlyValArg 726  
 Db 3489 TTACATTACATACCGCCACTAGAACCGAAGCAGCTCCAATAGCCACCGCGTGGTAT 3548  
 Qy 727 HisValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAsp 746  
 Db 3549 CACACCTCACCTCGCGGCCACCGCCATCTATTTCAATGGCTCTAATCTCGGATGAC 3608  
 Qy 747 GlnArgIleValSerTyrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspVal 766  
 Db 3609 CTGGCTGTGTCGAATATCTCTGACACCGCAGCAGCAGTTTGGCAGCGCGGCTCAT 3668  
 Qy 767 IleAspGlySerAspHisSer----- 773  
 Db 3669 GTGGCAGATACCGACAA-GGAGCCCGTAATGTTGTGAGTTAAGATTTAACTCTAAATCCC 3727  
 Qy 774 -----ValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyr 786  
 Db 3728 ACTCGCACTTATATTTATACAAATTTCCAGTTTGACCAATTTAGTACAAGGCGGCTAT 3787  
 Qy 787 ThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrVal 806  
 Db 3788 GTATTCACTGACTGTGAGCGATGATCAGGGTTTGACCAAGTTTGCAGATCTGTCTAGCGTT 3847  
 Qy 807 GluValGlnProAspProArgLysSerGlyLeuValGluLeuThrGlnValGlyVal 826  
 Db 3848 AATGTCGCGCGGATCCCAAGCTATTGAAATTTGGTTCAAAATGACCTGCCCATGGGATC 3907  
 Qy 827 GlyGlnLeuThrGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeuAsn 846  
 Db 3908 TCTGTGCTGTACAACTCCGAACCTGGATCGGTGTGCAAAACTGCGAGTTGTTCTGGGA 3967

Qy	847	ValLeuAspSerAspIleLysValGlnLysIleArg-----AlaHisSerAspLeuSer	864
Db	3968	GAT---GAGAAATAGATCCAGGTCAGGAGCTAAATATATACATTGCATCGGAT---GCT	4021
Qy	865	ThrValIleValPheTyrValGlnSerArgProPheLysValLeuLysAlaAlaGlu	884
Db	4022	ACTGTCTGTGCTTCTACGTAACAGATGGACAGGTT---AAGGCATTGGATGGTTTGC	4078
Qy	885	ValAlaArgAsnLeuHisMetArgLeuSerLysGlnLysAlaAspPheLeuLeuPheLys	904
Db	4079	GTGGAGCGGAGTGTGAGAACCCAACTGCAGAGGATGCATCTACTATGGCGCATTTGCA	4138
Qy	905	ValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAsp	924
Db	4139	GTG---GATATTCCGACCTCTGTTTGGCAGAGTATGCTCCGCTCATGGCTCGTGTAA	4195
Qy	925	ProLeuThrLysArgCysIleCysSerHisLeuTrpMet-----	937
Db	4196	CCCATACCAGAGCTTCATATCGAGCGGTTCTGGATGCCATCGGCTGGCTACTTCTTT	4255
Qy	938	-----	940
Db	4256	AATAACCCAGGAGCCAAATTCGGGTAGTTTAATAATATATACCCCGCATTAGTCGCTA	4315
Qy	941	IleGlnArgTyrIleTrpAspGlyGluSerAsnCysGluTrpSerIlePheTyrValThr	960
Db	4316	ATTTCTTTGTACATTTTTTCA-----GACTGGTCCATATATATACGTGTTT	4359
Qy	961	ValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCys	980
Db	4360	GTTCGTGTGATGTGGCGTCGCTCTGCTATCTGGAGTATTTTGGGCGCATAGCCTGTGCC	4419
Qy	981	CysLysArgGlnLysArgThrLysIleArgLysLysThr---LysTyrThrIleLeuAsp	999
Db	4420	TGCAGACAATCGAATAAAGCCGCTCTCGTCAGAAAGTGCAGAAATATTCATTGATTGGT	4479
Qy	1000	AsnMetAspGlu-----GlnGluArgMetGluLeuArgProLys	1012
Db	4480	AATAAGGATGAAGAAGCAGCCAAATTGAAGTACCATATGCCATCTCTCTCGTAGCGAAG	4539
Qy	1013	-----TyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetVal	1028
Db	4540	GATTTAATTATCTACTATTATTTTCACTT---GCAGATTCTCGAAACACATCGCTG---	4592
Qy	1029	SerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArg	1048
Db	4593	-----ACGGAAATCAGACAGGATTCGATGTGCTCTTTTTCAGACCCCGCACCAAGTCCA	4646
Qy	1049	GlyAsnProLysValSerMetAsnGlySerIleArgAsnGly	1062
Db	4647	GGTCTGGCGAGCAACAAGTGCACAAATTTCTTACAGTTCATGGA	4688

```

; FILE REFERENCES: AM1010601
; CURRENT APPLICATION NUMBER: US/10/717,597
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US 60/459,782
; PRIOR FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US 60/427,982
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 4904
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68
; LENGTH: 126990
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-717-597-68

Alignment Scores:
Pred. No.: 1,748-23 Length: 126990
Score: 403.00 Matches: 93
Percent Similarity: 67.57% Conservatives: 7
Best Local Similarity: 62.84% Mismatches: 23
Query Match: 7.22% Indels: 26
DB: 19 Gaps: 4

```

```

; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 284
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-764-390-1

Alignment Scores:
Pred. No.:          3,77e-34          Length:          284
Score:             477.00           Matches:          94
Percent Similarity: 100.00%         Conservative:     0
Best Local Similarity: 100.00%      Mismatches:      0
Query Match:       8.55%           Indels:          0
DB:                20              Gaps:            0

US-10-764-390-3 (1-1072) x US-10-764-390-1 (1-284)

Qy      662  HiValArgGlyProSerAlaValGluMetGluAsnLeuAspLysAlaLeuAlaThrVal 681
      |||
Db      282  CAGTCAGAGGCCCCAGTCGAGTCGAGATGGAAATAATTGCAAGCAATAGCCACTGTG 223

Qy      682  ThrGlyLeuGlnValGlyThrTyHisPheArgLeuThrValLysAspGlnGlnGlyLeu 701
      |||
Db      222  ACTGGTCTCCAGGTGGGGACCTACCACTTCGGTTTGACAGTAAAGACCCAGCAGGAGCTG 163

Qy      702  SerSerThrSerThrLeuThrValAlaValLysLysGluAsnAsnSerProProArgAla 721
      |||
Db      162  AGCAGCAGCTCCACCCCTACTGTGGCTGTGAAGAAGGAAATAATATAGTCCTCCAGAGCC 103

Qy      722  ArgAlaGlyGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGlySer 741
      |||
Db      102  CGGGCTGGTGGCAGACATGTTCTTGCTTCCCAATAATTCATTACTTTGGATGTTCA 43

Qy      742  ArgSerThrAspAspGlnArgIleValSerTyrlLeuTrpIle 755
      |||
Db      42  AGGTCTACTATGATGACCAAGAAATGTGTCTTATCTGTGGATC 1

```

RESULT 10  
US-10-764-390-1/c  
Sequence 1, Application US/10764390  
Publication No. US20040214212A1  
GENERAL INFORMATION:  
APPLICANT: Agensys, Inc.  
APPLICANT: Raitano, Arthur B.  
APPLICANT: Jakobowitz, Aya  
APPLICANT: Challita-Bid, Pia M.  
APPLICANT: Ge, Wangmao  
APPLICANT: Favis, Mary  
APPLICANT: Steven B. Kanner  
APPLICANT: Juan J. Perez-Villar  
TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
TITLE OF INVENTION: Entitled 254p10b Useful in Treatment and Detection of  
TITLE OF INVENTION: Cancer  
FILE REFERENCE: 51158-20081.00  
CURRENT APPLICATION NUMBER: US/10/764,390  
CURRENT FILING DATE: 2004-01-23  
PRIOR APPLICATION NUMBER: US60/442,526  
PRIOR FILING DATE: 2003-01-24

US-10-764-390-3 (1-1072) x US-10-717-597-68 (1-126990)

QY 433 ProProValAlaValSerProGlnLeuGlnLeuThrLeuProLeuThrSerAla 452  
 Db 120394 CCTTCTATGCGCCAGCTGCTCAGAACTCCTGGGCTCAAGCGATCCTCCACCTCAGCT 120335  
 QY 453 LeuLeuAspGlySerGlnSerThrAspThrGluLeuValSerTyrHis-----TTP 470  
 Db 120334 -----TCCACAGTGTGGGATTACAGCGTGAG-TCACCTGCACCCAGCCTGG 120288  
 QY 471 GluGluLeuAsnGlyProPheIleGluGluYThrSerValAspSerProValLeuArg 490  
 Db 120287 CATATTCCT-----CCACTAGCGCTTCTTCTCTTTTGT--- 120252  
 QY 491 LeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGly 510  
 Db 120251 -----AGTTTGACTGTATACAGACTCGACCGGA 120225  
 QY 511 AlaThrAsnSerThrThrAlaAlaLeuLeuValAsnAlaValAspTyrProProVal 530  
 Db 120224 GCACCTAACTTACAACTGCACGCCCTAATAGTGAACAATGTGTGGACTACCCACCACTT 120165  
 QY 531 AlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGly 550  
 Db 120164 GCTAATGAGGACCAAAATACACCACTAATCTTGCCCAAACTCCATCCTTGAATGA 120105  
 QY 551 AsnGlnSerSerAspAspHisGlnIleValLeuTyrGluTyrSerLeuGlyProGlySer 570  
 Db 120104 AACACAGACGTCAGCATCACAGATTGCTCTATGAGTGGTCCCTGGGTCTGGGAGT 120045  
 QY 571 GluGlyValHisValValMetGln 578  
 Db 120044 GAGGGCAACATGTGTGTATGCAG 120021

RESULT 12  
 US-10-050-704-89  
 ; Sequence 89, Application US/10050704  
 ; Publication No. US200305042A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; TITLE OF INVENTION: 62 Human Secreted Proteins  
 ; FILE REFERENCE: PZ039P1  
 ; CURRENT APPLICATION NUMBER: US/10/050,704  
 ; CURRENT FILING DATE: 2002-01-18  
 ; PRIOR APPLICATION NUMBER: 09/684,524  
 ; PRIOR FILING DATE: 2000-10-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/08979  
 ; PRIOR FILING DATE: 2000-04-06  
 ; PRIOR APPLICATION NUMBER: 60/128,693  
 ; PRIOR FILING DATE: 1999-04-09  
 ; PRIOR APPLICATION NUMBER: 60/130,991  
 ; PRIOR FILING DATE: 1999-04-26  
 ; NUMBER OF SEQ ID NOS: 344  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 89  
 ; LENGTH: 1420  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-050-704-89

Alignment Scores:  
 Pred. No.: 8,33e-26 Length: 1420  
 Score: 397.00 Matches: 83  
 Percent Similarity: 54.50% Conservative: 20  
 Best Local Similarity: 43.92% Mismatches: 54  
 Query Match: 7.11% Indels: 32  
 DB: 14 Gaps: 4

US-10-764-390-3 (1-1072) x US-10-050-704-89 (1-1420)

QY 903 PheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHis 922  
 Db 120045 -----AGTTTGACTGTATACAGACTCGACCGGA 120225

US-10-764-390-3 (1-1072) x US-10-798-512-89 (1-1420)

QY 923 CysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAsnLeuLeuGln 942  
 Db 63 TGTGACTGTTTCCACCAACGCTGATCTGTGACCCCTTTTGGATGGAGAATTCATCATCAAG 132  
 QY 943 ArgTyrIleTrpAspGlyGluSerAsnAspGluTrpSerIlePheTyrValThrValLeu 962  
 Db 123 GTGCAGCTGAGGGATGGAGACAGCACTGTGATGGAGCGTGTATATGTTTATCATCTGCT 182  
 QY 963 AlaPheThrLeuLeuValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLys 982  
 Db 183 ACCTTTGTCATTTGTTGCTTCGGAATCCTGCTTGGACTGTGATCTGTTGTTGTAAG 242  
 QY 983 ArgGlnLysArgThrLysIleArgLysLysThrLysTyrThrIleLeuAspAsnMetAsp 1002  
 Db 243 AGGCCAAAA---GGAAAAACCAAGAGGAAAGCAAGTACAAGATCCTCGGATGCCACGAT 299  
 QY 1003 GluGlnGluArgMetGluLeuArgProLysTyr----- 1013  
 Db 300 ---CAGGAAGCCTGGAGCTGAGCCAACTCCCGACGAGTAAAGAAAAGAGATGTCA 356  
 QY 1013 ----- 1013  
 Db 357 CTGAGTGGCCTGAACCAAGCTCATGCTACTGAGATGAAGAACACAGCAGGAGACCCCA 416  
 QY 1014 GlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGluSerGluPhe 1033  
 Db 417 GGCATCAACAGAAAGGCGCTTTCTAAGTAGCAGCTGTATGCACTCCGAGTCAGAGCTG 476  
 QY 1034 AspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAsnProLysVal 1053  
 Db 477 GACAGCGAT---GATGCCATCTTTACATGGCCAGCAGCAGAGAGGCGCAAACTCCTGCAT 533  
 QY 1054 SerMetAsnGlySerIleArgAsnGly 1062  
 Db 534 GGTCAAGATGGCTCTGTATCCCAACGCGC 560

RESULT 13  
 US-10-798-512-89  
 ; Sequence 89, Application US/10798512  
 ; Publication No. US20040152164A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; TITLE OF INVENTION: 62 Human Secreted Proteins  
 ; FILE REFERENCE: PZ039P1  
 ; CURRENT APPLICATION NUMBER: US/10/798,512  
 ; CURRENT FILING DATE: 2004-03-12  
 ; PRIOR APPLICATION NUMBER: US/09/684,524  
 ; PRIOR FILING DATE: 2000-10-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/08979  
 ; PRIOR FILING DATE: 2000-04-06  
 ; PRIOR APPLICATION NUMBER: 60/128,693  
 ; PRIOR FILING DATE: 1999-04-09  
 ; PRIOR APPLICATION NUMBER: 60/130,991  
 ; PRIOR FILING DATE: 1999-04-26  
 ; NUMBER OF SEQ ID NOS: 344  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 89  
 ; LENGTH: 1420  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-798-512-89

Alignment Scores:  
 Pred. No.: 8,33e-26 Length: 1420  
 Score: 397.00 Matches: 83  
 Percent Similarity: 54.50% Conservative: 20  
 Best Local Similarity: 43.92% Mismatches: 54  
 Query Match: 7.11% Indels: 32  
 DB: 19 Gaps: 4

```

QY 903 PheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHis 922
Db 3 TTCAGAGCCTTGGAGTCAACACTGTACATGTTCAGCTGAACCTGTCGACCATCGCCAC 62
QY 923 CysAspProLeuThrLysArgCysLysCysSerHisLeuTrpMetGluAsnLeuGln 942
Db 63 TGTGACTCGTTCACCAACAGCTGTATCTGTGACCCCTTTTGGATGGAGAAATTCATCAAG 122
QY 943 ArgTyrIleTrpAspGlyCysSerAsnCysGluTrpSerIlePheTyrValThrValLeu 962
Db 123 GTGAGCTGAGGATGGAGACAGCACTGTGATGGAGCGTGTATATGTTATCATGCT 182
QY 963 AlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysLysCysCysLys 982
Db 183 ACCTTGTCTATGTTGTCCTTGGATCCTGTCTGTGACTGTGATCTGTGTTGTTAAG 242
QY 983 ArgGlnLysArgThrLysLysArgLysLysThrLysTyrThrIleLeuAspAsnMetAsp 1002
Db 243 AGGCAAAA--GGAAAAACCCAGAGAGAAAGCAAGTACAGATCTCTGATGCCACGGAT 299
QY 1003 GluGlnGluArgMetGluLeuArgProLysTyr----- 1013
Db 300 ---CAGAAAGCCTGGAGCTGAAGCCACCTCCCGAGCAGGTAAAGAAAAGAGATGTCA 356
QY 1013 ----- 1013
Db 357 CTGAGTGGCTGAACCAAGCTCATGATAGTGGAGATGAAGAACCCAGCAGAGACCCA 416
QY 1014 GlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGluSerGluPhe 1033
Db 417 GGCATCAACACAGAAAGCCCTTTTGTCTAAGTAGCAGCCTGATGCATCTCCGAGTCAGAGCTG 476
QY 1034 AspSerAspGluAspThrIlePheSerArgGluLysMetGluArgLysAsnProLysVal 1053
Db 477 GACAGCGAT--GATGCCATCTTTACATGGCCAGACCCGAGAGAGAGGCGCAAACTCTCTGCAT 533
QY 1054 SerMetAsnGlySerIleArgAsnGly 1062
Db 534 GGTCAAGTGGTCTGTATCCCAACGGC 560

RESULT 14
US-10-282-122A-13443
; Sequence 13443, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangshu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578

```

```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13443
; LENGTH: 3612
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13443

Alignment Scores:
Pred. No.: 5,46e-11 Length: 3612
Score: 241.50 Matches: 192
Percent Similarity: 35.27% Conservative: 106
Best Local Similarity: 22.72% Mismatches: 322
Query Match: 4.33% Indels: 227
DB: 17 Gaps: 40

US-10-764-390-3 (1-1072) x US-10-282-122A-13443 (1-3612)
QY 171 SerGlyLysGlnGluProArgGlySerAlaGluTyrThrAspTrpGlyLeuLeuProGly 190
Db 1329 TCAGACAGAGTGTACCCCT--GGACCTCGGCGCGTACGCTATCATGCTTTTACTTCAGGG 1385
QY 191 SerGlu-GlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThrGl 210
Db 1386 CCCAGCCGTCGCCATGATCTGTCG-----TCGACGCGGTGCGGGTGAT----- 1431
QY 210 nGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeuPr 230
Db 1432 -----GAACCTTTTATCCGTCGTCGCGCGCGCACACGCTCAACGACGCTTGC 1481
QY 230 oGluArgSerValLeuLeuProLeuProThrThrProSerSerGlyGluValLeuGlu-- 249
Db 1482 CACGTCGCTGCATGTTGCGCTGAAGGTTACCGACGCCCGCGGCGACGCTACTCGAANA 1541
QY 249 ----- 249
Db 1542 CCGTATCGTGGGGTGAGAAATTCGAATATGATCCCTTCCACCTACCGTTGCGATGCCAG 1601
QY 250 -LysGluLysAlaSerGlnLeuGlnGlnSerSerAsnSer----- 263
Db 1602 CTCGCCCGACGTGACCCGCTGAGTTTCGACATCGCGAATCGAAAAAAGACTACATGCT 1661
QY 264 -----SerGlyLys 266
Db 1662 CGACAGTCGCTTCGACGCGCAACAGTACGGCTATGTCGGCTCTATCCGGGTGAAGGGCC 1721
QY 266 sGluValLeuMetProSerHisSerLeuProProAlaSerLeuGluLeu----- 282
Db 1722 GCGCAACTGGTGGCTCCACGACGAGCCGATGCGTCGATCCAGTGCCTTCGAAAGAA 1781
QY 283 -----SerSerValThrValGl 288
Db 1782 TCAGATGGCAGACGGAGGTGTCGGCTTGGCGGCTATCGCGCTACCGCTTGTCTCAGG 1841
QY 288 uLysSerProValLeuThrValThrProGlySerThrGluHisSer----- 303
Db 1842 CCGCGCATGAGTGCAGCATTTGGCTGTGCCGCGGTGAGTACACACACATTCCGGCTGA 1901
QY 304 -----IleProThrProThrSerAlaAlaProSerGluSer---ThrProSe 319
Db 1902 ATACCTGGCGCGCAATCCCAATCTCTCCGGGTACTTACGATGGGTTTTTGGCCTT 1961
QY 319 rGluLeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAl 339

```



```
Db 1962 GCAGCCAGAGGAGCAGCGCCAGCAGAGAAATCTCCAGAGGTCCATCTGAAGGT 2021
Qy 339 aGlyAspAsnLeuIleThrLeuProAspAsnGluValGluLeuIleValAlaPheValAl 359
Db 2022 GCGAGGTGAT-----GGCTGCCCGTAGTAATGTGGCGAGAGCGCGGTACAGT 2072
Qy 359 aProAlaProValGluThrThrTyraSerThrGluThrAsnLeuIleSerHisProTh 379
Db 2073 CGTCGCGAGCAGCGCTTGAACAGACTATGCGCTGGACGGGAGCGGCGAGCCGCCG 2132
Qy 379 rAsp-----TyrGlnGlyGluIleGlyGlnGlyHisGlnThrLeuAsnLe 395
Db 2133 GGACGAAAGCTGACCTATAAATGGCTGGTGGCAGCGGATCGTTCGCATCCCAATGC 2192
Qy 395 uSerGlnLeuSerValGlyLeuTyValPheLeysValThrValSerSerGluAsnAlaPh 415
Db 2193 GGATCAGGCCAAGCCAGCGCGGTGTCGCGAAG-----AATGCGAC 2234
Qy 415 eGlyGluGlyPheValAsnValThrVal-----LysProAlaArgA-gVa 430
Db 2235 GGGTGAGGGCGCTCATGAGCTGACCGCTCAGAACGAGTGGGCAAGACGGCGAAGCGAC 2294
Qy 430 lAsnLeuProValAlaValSerProGlnLeuGlnGluLeuThrLeuProLeuTh 450
Db 2295 G-----GTCAGGTACCGCGGTGGCGCCAGTCTC----- 2325
Qy 450 rSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSer----- 467
Db 2326 ---ACGATCAGCGCGGTAGCGCAGCAGGTGAAGCAACAGCAGCAGCGAGCTACAGCGC 2381
Qy 468 -----TyrHisTirp-----G1 471
Db 2382 GAATCGCAACTTCGACGCGCGGACCTACTCTGGAAGTGTCCAGCGTGGCGGCAAGCA 2441
Qy 471 uGluIleAsnGlyProPheIleGluGlyLeuThrSerValAspSerProValLeuArgLe 491
Db 2442 GCGCGGACGCGGTAGCGGGCGCAATGTGAACACCGCGTG----- 2481
Qy 491 uSerAsnLeuAspProGlyAsnTySerPheArgLeuThrValThrAspSerAspGly-- 510
Db 2482 ---GGTCTGATCGCGGCGAGTACACGGTGTCTGACGACGTAACGCGCGCGCGCA 2537
Qy 511 -AlaThrAsnSerThrThrAlaAlaLeuIleValAsnAsnAlaValAspTyPro---Pr 529
Db 2538 CCGGAGCGGAGCGCGCAGCAAACTGACCGTTGAGCGCAGCGTAGCGCTTCCGCGACC 2597
Qy 529 oValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsn-----SerIl 546
Db 2598 GGAGGCCAACCGCGGCGCGGACCGCGTGGTGGCGACGAACGACACCGCGGTTTGGCTA 2657
Qy 546 eThrLeuAsnGlyAsnGlnSerSerAspAspHis---GlnIleValLeuTyGluTrpSe 565
Db 2658 TCCGCTGTCGCGAGCGCAAGTCGCGACCCGCAACCGGCGGCGGATCAGTATCAATGGCG 2717
Qy 565 rLeu-----GlyProGlySerGluGlyLeysHisValVa 576
Db 2718 CGTGTGTGCGGCGCGGTTTCAGCTTGGCGCAACGGTGTGATCGTCCGATGGGAGCGATCGT 2777
Qy 576 lMetGlnGlyValGlnThrProTyLeuHisLeuSerAlaMetGlnGluGlyAspTyTh 596
Db 2778 ACCGAAGGCGAC-----ACCGCGACGGCGGT 2804
Qy 596 rPheGlnLeuIleValThrAspSerSerArgGlnGlnSerThrAla***-----ValTh 614
Db 2805 GTTCGAGCTGAAGTGAGCAACCGCGCGGCAAGCAGCGCGGCGGACCAAGAGGTGAC 2864
Qy 614 rValIleValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLysGl 634
Db 2865 CGCGATCGCG-----CCGCGCGTGACGCTCAGCGGC----- 2895
Qy 634 uLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerSerSerAsp----- 651
```

```
Db 2896 -----AGCAGCACCGTGCACGGCAGCGGGGCGAGCCAGCTACAAGGC 2936
Qy 652 -----AspHisGlyIleValPheTyHisTirp-----GluHisValArgGlyProSe 667
Db 2937 GAACCGCAACTTCGAGGAAGCGACCTATAGCTGCGACGCTCACCGATACGTCGGCGCAGCA 2996
Qy 667 rAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGl 687
Db 2997 GGTGTATCGGCGAGCGGTGCGGATTCGAACTGTCTC-----GCTACGTGAAGAAGCG 3050
Qy 687 yThrTyHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThLe 707
Db 3051 CAATACACCTCAAGTGTCTCGCATCGATCGATCAAGGGCGAGCGTAGC---GCGACGGC 3107
Qy 707 uThrValAlaValLysLysGluAsnAsnSerProProArg-----AlaArgAlaGl 724
Db 3108 GACCTCGCGCTGAAGGTGGAGCGCGGTGGCGCCCAAGCCGCTGGAGGCCAAATCAG 3167
Qy 724 yGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerTh 744
Db 3168 CGGT---CCGCTTGAATCAGGGCGGCGACCATCTTACTTGTGACCGTCCGAGCGCAATCGT 3224
Qy 744 rAspAspGlnArgIle---ValSerTyLeuTirpIleArgAspGlyGlnSerProAlaAl 763
Db 3225 CAACCCGAAACGGTACGCGGACCTTACTTGTGACCGTCCGCGCGCAG----- 3273
Qy 763 aGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThr-----As 780
Db 3274 -----CTGGCTCGAATGCCGAAACCGCGAGCGTCCGTTGACGGCGCTGCCAA 3323
Qy 780 nLeuValGluGlyVal---TyrThrPheHisLeuArgValThrAspSerGlnGly---Al 798
Db 3324 TGCCACAGTGGCGCGCAGTACCATCAAGCTCTCGGTGTACGACGACGAGAACAGCT 3383
Qy 798 aSerAspThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuVa 818
Db 3384 GTCCAGTACGCGGAGCTACACGGTGCAGCGTGAAAGCGCGGAAACACCGCGCGCCAGC 3443
Qy 818 lGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAspThrLeuVa 838
Db 3444 CTATCCGCGCTACAAGCGCGT-----ACGAGGTACAAGCGTGTGTAGCTGCT 3491
Qy 838 lArgGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLysValGlnLysIleAr 858
Db 3492 GCT-CAACTACGCTCAGCTCTATCAGTG-----CAAGCCGTTCCCGTACCGCGCTGCT 3544
Qy 858 gAlaHisSerAspLeuSerThrValIleValPheTy-ValGlnSerArgProPheLy 878
Db 3545 -GCA-----GTACGGCGCGCGCGCTATGAG 3570
Qy 878 sValLeuLysAla 882
Db 3571 CCGGTTAAAGCT 3583

RESULT 15
US-10-278-698-1034
; Sequence 1034, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmueller, Bruno
; APPLICANT: Haupt, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: 030027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1034
; LENGTH: 82027
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-1034
```

Alignment Scores:			
Pred. No.:	1.74e-07	Length:	82027
Score:	225.00	Matches:	252
Percent Similarity:	33.64%	Conservative:	186
Best Local Similarity:	19.35%	Mismatches:	416
Query Match:	4.03%	Indels:	448
DB:	21	Gaps:	69
US-10-764-390-3 (1-1072) x US-10-278-698-1034 (1-82027)			
QY	78	TyrLeuValSerCysProHisLysGluAsnCys---	GluProLysLysMetGlyProIle 96
DB	41174	TTCCAGGGTCAGTCCGCAAAATAAATATGTTGAGGGAAGGCTGAAATCGGAGCCAATT	41233
QY	97	ArgSerTyrLeuThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyr	116
DB	41234	GTTGGAGACATCCCAATTGATGTGCTGATGCTCCGCCACCTCCCAATATTGTGGAT---	41290
QY	117	GlyAspMetLeuAsnArgGlySerProSerGlyIleTyrGlyAspSerProGluAsp	136
DB	41291	-----GTGACACAGATTCAGTATCTCTAACTTGACACTGACCCC-----	41329
QY	137	IleArgLysAspLeu***PheLeuGlyLysAspTyrGlyLeuGluGluMetSerGluTyr	156
DB	41330	-----AAGAAAACCTGGTGGTCTCCCAATTACAGGGTAT	41362
QY	157	***AspAspTyrArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluPro	176
DB	41363	CATCTCGAGTTCAAGGAA---AGAAACAGCCCTTTTGTGGAAGAGAGCTAAACAAGACTCGG	41419
QY	177	ArgGlySerAlaGluTyrThrAspTyrGlyLeuLeuProGlySerGluGlyAlaPheAsn	196
DB	41420	ATAAGATCAGAGACTTTAAAGTCAGAGATTAACTGAAGTCTTGAATATGAATCCGA	41479
QY	197	-----SerSerValGlyAspSerProAlaValProAlaGlu-----	208
DB	41480	GTTATGGCAATCAATTTAGCAGGTGTGGC---AAGCCAAAGCTACCATCAGAGCCTGTT	41536
QY	209	ThrGlnGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLys	228
DB	41537	GTGGCACTGGACCCA-----ATTGATCCTCTCGA	41566
QY	229	LeuProGlu-----ArgSerValLeuLeuProLeuProThrThrProSer	243
DB	41567	AAACCTGAGTTATTAAACATAACAGGAATTGAGTACTCTCATTTGGACTGAACCTAAA	41626
QY	244	-----SerGlyGluValLeuGluLysGlu-----	251
DB	41627	TATGACGGTGGTCATAAGTTAACTGGATATATAGTGGAGAAGCAGATCTACTCTCGAAG	41686
QY	252	-----LysAlaSer-----	254
DB	41687	TCTTGGATGAAGCCCAACCATGTTAATGTCCCAGAAATGTCCTTTACTGTAACGACCTT	41746
QY	255	-----GlnLeuGlnGluGlnSerSerAsnSerSerGlyLysGluVal	268
DB	41747	GTTGAGGGTGGAATAATATGAATTCAGAAATTAGAGCAAGAAATACAGCAGGT---GCTATC	41803
QY	269	LeuMetProSerHisSerLeuProPro-----AlaSerLeuGluLeuSerSerValThr	286
DB	41804	AGTGCTCCATCAGAAAGTACAGAAACCATTAATTTGCAAGGATGAATACGAGGCACCAACA	41863
QY	287	ValGluLysSerProVal-----LeuThrValThrProGlySerThr-----	300
DB	41864	ATTGTCTTTGATCCCAATAAAGATGGGCTAACAAATTAAGCAGGGGATACCAATTGTT	41923
QY	301	-----GluHisSerIleProThrProProThrSerAlaAlaPro	313
DB	41924	TTGAATGCCATTAGCAATTTCTGGCAAAACCCCTTCCAAATCAAGTTGGTCCAGGAGGA	41983
QY	314	SerGluSerThrProSer-----GluLeuProIleSerProThrThrAlaProArg	330
DB	43004	GAGCCAGCGGTGATTTCCAGATGTTTGAATTCGTAGAACCGGAGATGGCTCTCTGATTTTGA	43063
Db	41984	AAAGACATTAGACCATCAGATATCACTCAGATAACTTTCAACCCCAACATCTTCCATGCTT	42043
QY	331	ThrValLys-----GluLeuThrValSerAlaGly-----	340
DB	42044	ACTATCAAGTATGCCACTAGAAAGATCGGGTGAATATACATCCTACCTGCTACCAATCCT	42103
QY	341	-----AspAsnLeuIleThrLeu-----ProAspAsn	350
DB	42104	TTTGGCAGGAAGTGGAAACATGTGAAGTAACTCCTTGTGATACCTTGGTCTCCCGCAGGT	42163
QY	351	GluValGlnLeu-----LysAlaPheValAlaProAlaProProVal	364
DB	42164	CCTGTTGAAATCAGTAATGTTTCTGCTGAAAGCAACACTTACATGAGACCTCCCTTG	42223
QY	365	Glu-----ThrThrTyrAsnTyr	370
DB	42224	GAAGATGGCGCTCACCATTAACTCTATATATCTTGAAGAGAGAAACAGCGGACTT	42283
QY	371	GluTyrAsnLeuIleSerHisProThrAspTyrGlnGly-----Glu	384
DB	42284	TTGTGGACAGTGGTTTCTGAA-----GATATTCACTCTTCAGGCATGTGGCAACCAA	42337
QY	385	IleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrVal	404
DB	42338	CTTATCCAAAGGAATGAG-----TACATC	42361
QY	405	PheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrVal	424
DB	42362	TTCCGG-----GTCTCAGCTGTAAACCACTATGGCAAGGA---GAACCTGTACAGTCT	42412
QY	425	LysProAlaArgArgValAsnLeu-----ProProValAlaValValSerProGlnGlu	442
DB	42413	GAACCTGTCAAAATGTAGACAGATTTGGTCCCTCGCCCTCTGAAAAACAGAGGTA	42472
QY	443	GlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspAsp	462
DB	42473	TCAAATGTCACT-----AAGAACACTGCCACTGCAGCTGGAAGGCCAGTGGATGAT	42526
QY	463	-----ThrGluIleValSerTyrHisTyrGluGluIleAsnGlyProPheIleGluGlu	480
DB	42527	GGTGGCAGCAAAATTACAGGATATCATGTAGAAAGGAGAGAAAGAAAGCCTCGCATGG	42586
QY	481	LysThrSerValAspSerProVal-----LeuArgLeuSerAsnLeuAspPro	496
DB	42587	GTGAGACCAATAAAACACACAGTTCCGATCTCAGGTGCAAAAGTAACAGGACTGCAAGAA	42646
QY	497	GlyAsn-----TyrSerPheArgLeuThrValThrAsp-----SerAspGlyAlaThrAsn	513
DB	42647	GGAAGCACTACGAATTCGTGTGTCAGTCAGAAACACAGAGCAAGAAATGGTCCACCCAGT	42706
QY	514	SerThrThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProProValAlaAsnAla	533
DB	42707	GAGGCTTCAGATTTCTGTATGAAGATGCAGCA---TATCTTCCAGGACCACCTTCA	42763
QY	534	GlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSer	553
DB	42764	AATCGCATGTCACGTACTACCAAGAAATCTGCTTCTTGGCATGGCGCAAGCCTCAT	42823
QY	554	SerAsp-----AspHisGlnIleValLeuTyrGlu	563
DB	42824	TATGATGTGGACTTGAATCACTGGCTATGTCTGGAGCATCAAAAAAGTAGGAGACGAG	42883
QY	564	---TTP-----	564
DB	42884	GCCTGGATAAAGATACACAGGAACCGCCCTCAGAAATCACTCAGTTGTTGTTCTGAT	42943
QY	565	-----SerLeuGlyProGlySerGlyGly	572
DB	42944	CTTCAGACTAAGAAAAATACAACTTCAGAAATCAGTGCATCAACAGATGCAGGTGTGG	43003
QY	573	LysHisValValMetGlnGlyValGln-----ThrProTyrLeuHis	586
DB	43004	GAGCCAGCGGTGATTTCCAGATGTTTGAATTCGTAGAACCGGAGATGGCTCTCTGATTTTGA	43063

```
Qy 587 LeuSerAla-----MetGlnGluGly----- 593
Db 43064 CTAGATCCGAGCTTCGAAGAACAACCTTGTGTGTAGACGAGACTCAGTATTAGGATATTT 43123
Qy 594 -----AspTyrThrPhe-----GlnLeuLysVal 601
Db 43124 GTGCCAATTAAAGTCGTCCTGCTGCTGAGTGCATGAGCAAGCAAGATAACATCAACCTG 43183
Qy 602 ThrAspSerSerArgGlnSerThrAla***ValThrValIleValGlnProGluAsn 621
Db 43184 AAAAACCGAGCAACATTGAAATAACGGAATCAATTACTTCTCTGATTATCCAGAAATGT 43243
Qy 622 AsnArg-----ProProValAlaValAla 629
Db 43244 AACAGATATGATACCGGTAAATTTGTCTATGACCAATTTGAAACCCGCTGGGAAGAAAGT 43303
Qy 630 GlyProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSer 649
Db 43304 GGCCTTTGAACGTACGAGTCTTGACACGCGCCAGTCCCTCAACCTCGCGCTTACA 43363
Qy 650 SerAspAspHisGlyIleValPheTyrHisTrpGlu-----HisValArgGlyProSer 667
Db 43364 GACATCAAAAGGACAGTGTCACTCGCTGCGACCTCCCTCTGATAGATGGAGGTCA 43423
Qy 668 AlaValGluMetGluAsnIleAspLysAlaIleAlaThr----- 680
Db 43424 CGTATAACAACTACATTGTAGAGAAACGTGAAGCAACACGCGAAATCTTATTCACAGCC 43483
Qy 681 -----ValThrGlyLeuGlnValGly---ThrTyr 689
Db 43484 ACCACTAAGTCCATAATGACACATATAAAGTTACCGCTTGTCTGAAGGTGTGAATAT 43543
Qy 690 HisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrVal 709
Db 43544 TTCTTCAGAGTGTGCGACAGAAATGAATATGGAATTTGGTGAGCCCAACAGAACTACAGAG 43603
Qy 710 AlaValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyArgHisValLeu 729
Db 43604 CCCGTAAAGCTCTGAAGCACCATCTCCA----- 43633
Qy 730 ValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIle 749
Db 43634 -----CCAGACACGCTTAACATCATGGACATAACTAAGACACC----- 43672
Qy 750 ValSerTyrLeuTrpIleArg-----AspGlyGlnSerProAlaAlaGlyAspVal 766
Db 43673 GTACGCTGGCATGGCTTAAGCCCAACACAGATGTTGGCAGCAAGATCACTGGCTATGTG 43732
Qy 767 IleAsp-----GlySerAsp-----HisSerValAlaLeuGln----- 777
Db 43733 ATTGAAGCCCAAGAAAGAGCTCTGACCACTGGACCCACATCACAAACCGTGAAGGGTTA 43792
Qy 778 -----LeuThrAsnLeuValGluGly---ValTyrThrPheHisLeuArgValThr 793
Db 43793 GAAATGTGTGTGGAAGATCTAACTGAAGGAGAGGAATATACCTCCAAAGTGTATGGCAGTG 43852
Qy 794 AspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnPro-----Asp 811
Db 43853 AACAGCGCGGGAGAGT-----GCCCTAGAGAAAGCAGACCCGCTCATTTGTC 43900
Qy 812 ProArgLysSerGlyLeuValGluLeuThrLeuValGlyValGlyGln----- 828
Db 43901 AAGGACGACAGACAATGCTTCCAGAGCTGGATCTCGT---GGCATCTATCAGAAATCGGTC 43957
Qy 828 ----- 828
Db 43958 ATTGCCAAAGCTGGTGACAAACATCAAGTTGAAATTCAGTCTCGGTGCGACCGAAGCCC 44017
Qy 829 ---LeuThrGluGlnArgLysAspThrLeuValArgGln----- 840
Db 44018 ACAGTGACATCGAANAAGGAGACCAAAATTTCTTAACACACAGACAGAGATTAAATTTGAA 44077
```

```
Qy 841 -----LeuAlaValLeuLeuAsnVal-----LeuAspSerAspIleLysVal 854
Db 44078 ACCACAGCGACTTCAACCACTTTAAATATCAATGAGTGTGCAAGAGTATGGGGCCC 44137
Qy 855 GlnLysIleArgAlaHisSer-----AspLeuSerThrValIleValPheTyrVal 871
Db 44138 TATCCATTAAACAGCAAGGAACATTGTAGGAGAGTGTGTGATGTCATCACCATCAAGTC 44197
Qy 872 GlnSer-----ArgProPhePheLysValLeuLysAlaAlaGluValAlaArgAsnLeu 889
Db 44198 CATGATATCCAGGGCCACCTACTTGGCAATCAATCAATTTGTATGAAGTTTCA----- 44248
Qy 890 HisMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAsp 909
Db 44249 -----TCTGATTTTGTAACTTCTCT----- 44269
Qy 910 ThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArg 929
Db 44269 ----- 44269
Qy 930 CysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGlu 949
Db 44270 -----TGGGACCCCACT 44281
Qy 950 SerAsn-----CysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeu 966
Db 44282 GAGAACGATGTTGTGTACCAATGAACAATATATGTAGTGAATGCGGCAG----- 44332
Qy 967 IleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLysArg 986
Db 44333 -----ACTGACAGTACTACCTGGGTGAGTTAGCA-----ACC 44365
Qy 987 ThrLysIleArgLysLysThrLysTyrThr---IleLeuAspAsnMetAspGluGlnGlu 1005
Db 44366 ACCGTTATACGTACTACTATAAAACCCACCCGCTTACTACTGGATTAGATCAGTTC 44425
Qy 1006 ArgMetGluLeuArgProLysTyrGlyIleLysHisArgSerThrGluHisAsnSerSer 1025
Db 44426 CGTGTAAAGCTCAGATAGATATGGAGTTGGACCCAGGCATCACA-----TCAGCA 44476
Qy 1026 LeuMetValSerGluSerGluPhe----- 1033
Db 44477 TGCATAGTTGCCAACTATCCATTTAAGTTCTCGACCTCTGCTGCTACCCCTCAGGTAAC 44536
Qy 1034 -----AspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAsnPro 1051
Db 44537 GCAGTTACCAAGGATTCAATGACAATAGCTGGCATGAGCCACTTTCTGTGTTGGTGAAGC 44596
Qy 1052 LysVal-----SerMetAsnGlySerIleArgAsnGlyAlaSerPheSerTyrCys 1068
Db 44597 CCCATTTTAGGATATCATGTTTGAAGAAAGAAAGACGAAATGTGTTCTCTGCGACACTGTG 44656
Qy 1069 SerLys 1070
Db 44657 AGCAAA 44662
```

Search completed: October 12, 2005, 16:49:57  
Job time : 4109 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 12, 2005, 10:31:52 ; Search time 414 Seconds

(without alignments)  
4236.929 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSILLLVITAGC.....VSMNGSIRNGASRYSCKDR 1072

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool p/US10764390/runat 12102005 110204 22353/app query.fasta\_1.1223  
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10764390.cgn2\_1 1105 @runat 12102005 110204 22353 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					
C 1	443	7.9	553	4	US-09-270-767-827
C 2	443	7.9	553	4	Sequence 827, App
C 3	322	5.8	3666	4	US-09-270-767-16109
C 4	317	5.7	4287	4	Sequence 16109, A
C 5	317	5.7	1584	4	Sequence 484, App
C 6	211.5	4.9	3282	4	Sequence 2590, App
C 7	215.5	3.9	2481	4	Sequence 1119, App
C 8	207.5	3.7	1119	4	Sequence 6552, App
C 9	207.5	3.7	10838	4	US-09-902-540-6552
C 10	201.5	3.6	25165	3	Sequence 3204, App
C 11	201	3.6	1684976	4	US-09-248-796A-3204
C 12	201	3.6	1664976	4	Sequence 2872, App
					Sequence 1013, App
					Sequence 39, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 1, Appl

13	199	3.6	320	4	US-09-621-976-13082	Sequence 13082, A
14	197.5	3.5	16047	4	US-09-902-540-1136	Sequence 1136, Ap
15	196	3.5	15512	2	US-08-853-659A-5	Sequence 5, Appl
16	196	3.5	15512	2	US-08-853-659A-8	Sequence 8, Appl
17	196	3.5	15512	2	US-08-853-659A-63	Sequence 63, Appl
18	196	3.5	15512	2	US-08-853-659A-66	Sequence 66, Appl
19	196	3.5	15512	2	US-08-853-659A-2	Sequence 2, Appl
20	196	3.5	24701	2	US-08-853-659A-3	Sequence 3, Appl
21	196	3.5	24701	2	US-08-853-659A-60	Sequence 60, Appl
22	196	3.5	24701	2	US-08-853-659A-61	Sequence 61, Appl
23	195	3.5	14255	1	US-08-320-559-1	Sequence 1, Appl
24	195	3.5	14255	1	US-08-327-392-1	Sequence 1, Appl
25	195	3.5	14255	1	US-08-306-691B-55	Sequence 55, Appl
26	195	3.5	14255	3	US-08-545-860D-1	Sequence 1, Appl
27	195	3.5	14255	5	PCT-US94-04496-1	Sequence 1, Appl
28	194	3.5	11907	3	US-08-061-376-4	Sequence 4, Appl
29	193	3.5	6763	4	US-09-949-016-3399	Sequence 3399, Ap
30	192.5	3.4	17592	4	US-09-902-540-1138	Sequence 1138, Ap
31	191.5	3.4	4063	4	US-09-902-540-595	Sequence 595, App
32	185.5	3.3	2643	4	US-09-902-540-2841	Sequence 2841, Ap
33	183.5	3.3	8211	4	US-09-252-991A-13656	Sequence 13656, A
34	183.5	3.3	13807	3	US-09-052-469-5	Sequence 5, Appl
35	183.5	3.3	13807	4	US-08-422-582-5	Sequence 5, Appl
36	183.5	3.3	13807	4	US-09-052-262-5	Sequence 7, Appl
37	183.5	3.3	14148	3	US-09-052-469-7	Sequence 7, Appl
38	183.5	3.3	14148	4	US-08-422-582-7	Sequence 7, Appl
39	183.5	3.3	14148	4	US-09-052-262-7	Sequence 80, Appl
40	181	3.2	9164	4	US-09-814-915A-80	Sequence 12384, A
41	180	3.2	39154	4	US-09-949-016-12384	Sequence 12384, A
42	180	3.2	39154	4	US-09-949-016-12801	Sequence 12801, A
43	180	3.2	39443	4	US-09-949-016-14326	Sequence 14326, A
44	180	3.2	39443	4	US-09-949-016-14327	Sequence 14327, A
45	179.5	3.2	13941	4	US-09-799-451-341	Sequence 341, App

ALIGNMENTS

RESULT 1  
US-09-270-767-827/c  
; Sequence 827, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 827  
; LENGTH: 553  
; TYPE: DNA  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-827

Alignment Scores:  
Pred. No.: 5.29e-32 Length: 553  
Score: 443.00 Matches: 92  
Percent Similarity: 62.30% Conservative: 22  
Best Local Similarity: 50.27% Mismatches: 69  
Query Match: 7.94% Indels: 0  
DB: 4 Gaps: 0

US-10-764-390-3 (1-1072) x US-09-270-767-827 (1-553)

Qy	473	IleasnglyProPheIleGluLysThrSerValaaspSerProValaarguensr	492
		:	:
Db	551	ATTTGGGACCGTGGTTACCAACCGATTTTGGCAGAGTTAACACCTCCAGTTGAT	492
		:	:
Qy	493	AsnLeuaspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThr	512
		:	:
Db	491	CTCATCGCGCGGAGACTACACATTCAAGTACCGTACCGATTCAATACCGTACC	432
		:	:

```

Qy 513 AenSerThrAlaAlaLeuIleValAenAenAlaValAspTyrProProValAlaAen 532
Db 431 AATTCAACCACTGCCACGATACGAGTTCTTAAGGAGACCGATTATGCTCCGGTGGCAAT 372
Qy 533 AlaGlyProAsnHisThrIleThrLeuProGlnAenSerIleThrLeuAenGlyAenGln 552
Db 371 GCCGCGGATGCGGTGATCTTGTATTGTCACCAACAATAATGTCACTCTGAATGGCACAGCC 312
Qy 553 SerSerAspAspHisGlnIleValLeuTyrGluThrProSerLeuGlyProGlySerGluGly 572
Db 311 AGTTCGGATGATCAGAGATCTTGTCTGGAGTGGACCAAGGATGCTAGCGACAGGCC 252
Qy 573 LysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGlu 592
Db 251 AAGCTGTGGATATGCAGATACAGAACACCCCTATGTTCACTGTCCAATTGGAGGAG 192
Qy 593 GlyAspTyrThrPheGlnLeuIleValThrAspSerSerArgGlnGlnSerThrAla*** 612
Db 191 GGCATGTACACTTTTGTCTTAAAGTAAACCGATGGCGAGTGGCAATCGAGTACAGCTAAG 132
Qy 613 ValThrValIleValGlnProGluAenAenArgProProValAlaValAlaGlyProAsp 632
Db 131 GTCCATGTGTTTGTGAAGCTCCCAAGAAATCTCCACCAAGTTCCTGAGCGAGGAAGTAAT 72
Qy 633 LysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerAspAsp 652
Db 71 ACGACTACTAGTTCGCCATTAAATGGGTTCTTTGAATGGCTCCGATTCGAAGACGAC 12
Qy 653 HisGlyIle 655
Db 11 ATTGGCATC 3

```

## RESULT 2

```

US-09-270-767-16109/c
; Sequence 16109, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16109
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16109

```

```

Alignment Scores:
Pred. No.: 5.29e-32 Length: 553
Score: 443.00 Matches: 92
Percent Similarity: 62.30% Conservative: 22
Best Local Similarity: 50.27% Mismatches: 69
Query Match: 7.94% Indels: 0
DB: Gaps: 4

```

US-10-764-390-3 (1-1072) x US-09-270-767-16109 (1-553)

```

Qy 473 IleAsnGlyProPheIleGluGluLysThrSerValAspSerProValLeuArgLeuSer 492
Db 551 AYTTCGGACCGATTGGTTACCAACCCAGTTTCCAGAGGTTAAACACCTTCAGTTGGAT 492
Qy 493 AsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerGlyAlaThr 512
Db 491 CTCACATCGCGGGGAATACACATTCAAGCTAACCGTAAACCGATTCAAATAACGCGACC 432
Qy 513 AsnSerThrThrAlaLeuIleValAenAenAlaValAspTyrProProValAlaAen 532
Db 431 AATTCAACCACTGCCACGATAGCAGTTCTTAAGGAGACCGATTATGCTCCGGTGGCAAT 372
Qy 533 AlaGlyProAsnHisThrIleThrLeuProGlnAenSerIleThrLeuAenGlyAenGln 552

```

```

Db 371 GCCGCGGATGCGGTGATCTTGTATTGTCACCAACAATAATGTCACTCTGAATGGCACAGCC 312
Qy 553 SerSerAspAspHisGlnIleValLeuTyrGluThrProSerLeuGlyProGlySerGluGly 572
Db 311 AGTTCGGATGATCAGAGATCTTGTCTGGAGTGGACCAAGGATGCTAGCGACAGGCC 252
Qy 573 LysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGlu 592
Db 251 AAGCTGTGGATATGCAGATACAGAACACCCCTATGTTCACTGTCCAATTGGAGGAG 192
Qy 593 GlyAspTyrThrPheGlnLeuIleValThrAspSerSerArgGlnGlnSerThrAla*** 612
Db 191 GGCATGTACACTTTTGTCTTAAAGTAAACCGATGGCGAGTGGCAATCGAGTACAGCTAAG 132
Qy 613 ValThrValIleValGlnProGluAenAenArgProProValAlaValAlaGlyProAsp 632
Db 131 GTCCATGTGTTTGTGAAGCTCCCAAGAAATCTCCACCAAGTTCCTGAGCGAGGAAGTAAT 72
Qy 633 LysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerAspAsp 652
Db 71 ACGACTACTAGTTCGCCATTAAATGGGTTCTTTGAATGGCTCCGATTCGAAGACGAC 12
Qy 653 HisGlyIle 655
Db 11 ATTGGCATC 3

```

## RESULT 3

```

US-09-902-540-484
; Sequence 484, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 484
; LENGTH: 3666
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-484

```

```

Alignment Scores:
Pred. No.: 3.54e-19 Length: 3666
Score: 322.00 Matches: 231
Percent Similarity: 35.54% Conservative: 123
Best Local Similarity: 23.19% Mismatches: 343
Query Match: 5.77% Indels: 301
DB: Gaps: 54

```

US-10-764-390-3 (1-1072) x US-09-902-540-484 (1-3666)

```

Qy 16 ThrIleAlaGlyCysAlaArgLysGlnCysSerGluGlyArgThrTyrSerAsnAlaVal 35
Db 843 ACTACTACGGGTGCGCGCGCGCTTAGCCACCGCACCGCTGGCGTTCGAGCGCA 902
Qy 36 IleSer-----ProAsnLeuGluThrThrArg--- 44
Db 903 ACGCGCTGACCTTCGCGCACATCGAGAACGGCAATCCCTCCGACGAGCCACCGCTCG 962
Qy 45 -----IleMetArgValSerHisThrPheProValValAspCysThrAlaAlaCys--- 61
Db 963 CTACGGTACAGCGCGGCTCGGCAACTCGCGCG-----TGACACAGCGCGCGAGG 1013
Qy 62 CysAspLeuSerSerCysAspLeuAlaThrTrpPheGluGlyArgCysTyr----- 78

```



Db 1014 TGTGG-GCCTCC-----ATGCTCTGG-----GAGTGCTAGTGTGCTC 1051  
Qy 79 LeuValSerCysProHisLysGluAsnCyseGluProLysLysMetGlyProIleArgSer 98  
Db 1052 CTCAGGCCACCCCTTCCAGGAGGCCAGAGCCGATGAACACTACTGCTGGTCCGCCGG 1111  
Qy 99 TyrLeuThrPheValLeuArgPro----- 106  
Db 1112 TACAAGGCCACGCCGCTCTCCCCACCTTCCTCAGGGCGCGAGCGCTGCTCGCGGTC 1171  
Qy 107 -----ValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgGly 124  
Db 1172 ACCGCCGCGAGCACCGCGCGACTACACAGCGCTTCTCCAGGCTTCGCCCAAGCGGCT 1231  
Qy 125 SerProSerGlyIleIleTrpGly-----AspSerProGluAspIleArgLysAspLeu 141  
Db 1232 GCCGCATGGCGGAGGCGCGGACCGTGACTCCATGACACCAACGCGGGCGTGGAG 1291  
Qy 142 \*\*\*PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTyr\*\*\*AspAspTyrArg 161  
Db 1292 AGCTTCGCGCGGCGCAACTTC---ATCGAGTCACTCCCTGCGGCTCGATGACGCCCTG 1348  
Qy 162 Glu---LeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAla 180  
Db 1349 GAGGCTGCGACAGGAC-----GGCGTGCCTC 1375  
Qy 181 GluTyrThrAspTrpGlyLeuLeu-----ProGlySerGlu 192  
Db 1376 GACGTGGCGAGACGGGCTGCTGCGGTGAGCGTGCGCAACGTGGCGAGCGGTTCGCTG 1435  
Qy 193 GlyAlaPheAsn----- 196  
Db 1436 GCGCCCTTCTCCGCAACCGTCTCCACCACAGCAGCAGCGCCACGCTGGAGTTCCCTCG 1495  
Qy 197 -----SerSerValGly----- 200  
Db 1496 GGCACACCCCTGACCTTCAGTCTCATGGCGCGCGCGGCGCGCCACCGCCCTCCCG 1555  
Qy 201 -----AspSer 202  
Db 1556 GTGAGGTGGCGCGCGCGAGCCGACGACACACGCGCGCTGCGCATCCACTTCGACGAG 1615  
Qy 203 ProAlaValProAla-----GluThrGlnGlnAspProGluLeuHisTyr----- 217  
Db 1616 CCCTCGCTCCCGCGCAGCGCAGCAGCAGCGTTCGACCTCGGGTGAACACTAGCAGAA 1675  
Qy 218 LeuAsnGluSerAlaSer-ThrProAlaProLysLeuProGluArgSerValLeuLeuPr 237  
Db 1676 GTGCCGCGCGCGCGGAGGAGGACATTCGAGGGCGCCCTGTCATGGAACCTCGTCC 1735  
Qy 237 oLeuProThrThrProSer-----SerGlyG1 246  
Db 1736 AGTGGCTCAACCCCTACGGCGAGGCGGAGCTGGAAGTTCAGGGTCCGGAGGGCGGCGC 1795  
Qy 246 uValLeuGlyLysGluLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLys 266  
Db 1796 TACATGACGCGCGCCCAACCGCGCAGTACAGCGGACCTTCTCTCATCTCCCTGGATG 1855  
Qy 266 sGluValLeuMetProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValTh 286  
Db 1856 AAGTGAATGAGACCGCGG-----CCTTCTCGTTCAGCTTCGCTTACCGTACCGC 1900  
Qy 286 rValGluLysSerProValLeuThrValThrProGlySerThrGluHisSerIleProTh 306  
Db 1901 CACTCGTTCGAGTCCG-----ACATCTACCCCGCGGCT-----CCATC 1939  
Qy 306 rProProThrSerAlaAlaProSerGluSerThrProSerGluLeuProIleSerProTh 326  
Db 1940 TTCCCGTACTACGAGCGCGGCTCCTCGAGTTCACCCACG---ACGGCTCGCAGTGGTAC 1996  
Qy 326 rThrAlaPro-----ArgThrValLysG1 334  
Db 1997 GACATCGCGCGATGGGCGTGGCTTCCATCTACTCCATCCAGCCCGACGAGCGGCTCCCG 2056

Qy 334 uLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAsnGluValGluLe 354  
Db 2057 GGCTACATCAGCAGGGGC----- 2075  
Qy 354 uLysAlaPheValAlaProAlaPro-ProValGluThrThrTyrAsnTyrGluTrpAsnL 374  
Db 2076 -----CCACCCCGTGTATCGCAACGC----- 2100  
Qy 374 eulleSerHisProThrAspTyrGlnGlyGluIleLysGlnGly----- 388  
Db 2101 --GTGGGTGGCGCAGAACGTCACGACGAGAACCGCAAGACGGTGGCTTCCCGAGTGG 2158  
Qy 389 --HisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrVal---PheLysV 407  
Db 2159 TCCACCCCGCAG--CTGCGGCTGGCGGACATGTTCGCCGCAAGAACGTCAGTTCGCT 2216  
Qy 407 alThrValSerSerGluAsnAlaPheGlyGlu-----GlyP 419  
Db 2217 TCCGATTTGTACCGACACCGCGGTGGCGGTACGCTTCGACCTGGACACGTCGCT 2276  
Qy 419 heValAsnValThrValLysPro---AlaArgArgVal----- 430  
Db 2277 TCACCAAGTGGAGCGCATGCCCTTCAGCGCGCGGTGGACGAGCAGGGCGACGGCGCA 2336  
Qy 431 -----AsnLeuProProValAlaVal-----SerProG 441  
Db 2337 CCTGCAACCGACGCGCGTGGCCAAACCGCGCGCGACGTCGGCAGGCCCGCCCGG 2396  
Qy 441 lnLeuGlnGluLeuLeuLeuProLeuThrSerAlaLeu-----lleA 455  
Db 2397 TGTATGACGCGTGAGAAACCCCGAGACGGGGGACCTGGAGCTCGTGGCCACACGCTGG 2456  
Qy 455 spGlySerGlnSerThrAsp---AspThrGluIleValSerTyrHisTrpGluLeuLeA 474  
Db 2457 ACGGCACCGCAGCTATGACCGGAGGGCCAGCCCTCACCTACCTACCTGGACGAGGTGA 2516  
Qy 474 snGlyProPheIleGluGluLysThrSerValAspSerProValLeuArgLeuSerAsnL 494  
Db 2517 GCGGG-----CGGCGGTCACTCGTGTGAACG 2543  
Qy 494 euAspProGlyAsnTyrSerPheArgLeuThrValThrAspSer----- 508  
Db 2544 CGGACACGGCCACCCCGTCTGTTACGGCGGATGTCGCGACACTCATCTCTGAGTTCC 2603  
Qy 509 -----AspGlyAlaThrAsnSerThr-----ThrAlaAlaLeuIleV 521  
Db 2604 AGCTCGTGTGGCGATGGCTGGAGACCGCGGCGCACACAGGTCCGGGCTGGGTCC 2663  
Qy 521 alAsnAsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThrL 541  
Db 2664 TCAAC-----GTCAACCGCGCCCGCCACGCGGTGGTGAACG 2699  
Qy 541 euPro-----GlnAsnSerIleThrLeuAsnGlyAsnGlnInserS 554  
Db 2700 CGCGCGCGAGGTGGCGGAGTTCCTCCGAGCGCGGTGACGCTGGACCCCGCGCTCGA 2759  
Qy 554 erAsp---AspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyL 573  
Db 2760 GCGACCGCGAGCGGAGCTCATCTCTACAAGTGGCG-----CAGAGCTCCGCTC 2810  
Qy 573 ybHisValValMetGln-----GlyValGlnThrProTyrLeuHisLeuSerAlaMetG 591  
Db 2811 CGCGCGTGGCTCAAGCGGGCTCCAGCGCGCATCGCGGAGTTCGAAGTGGCCGAGGTGG 2870  
Qy 591 ln---GluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnInserT 610  
Db 2871 CGGTGATCTACTCATCTTCACCTTCAGCTGGTGGTGAACACTCGAGGACCGAGCAGA 2930  
Qy 610 hrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaG 630  
Db 2931 CGGCGACCGTGACGCTGATGTGGCGC---AACGTTGACCGGCGCGCTGGCGATGCGG 2987

QY 630 lyProAspGlyGluLeuPheProValGluSerAlaThrLeuAspGlySerSers 650  
 Db 2988 GC---TCCGAGCAGGTGGTGGACCGCTGGCCACCGCGCTCAGCGCGGTGGTG 3044  
 QY 650 exAspHisGlyLeuValPheTyrHis-TyrGluHisValArgGlyProSerAlaVal 669  
 Db 3045 GACCGGACGGGATGCCCTACCTGTAGTGGCTCAGGTGGAGGGCCCC---CGCGTG 3101  
 QY 670 GluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnVal----- 686  
 Db 3102 GAGTGCAGGGAGCGCCACGCGCCACCCCGCTTCACCGCCCCCGGGTGGCGCGGAG 3161  
 QY 687 GlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThr-----Ser 705  
 Db 3162 ACGGTGCTGGCTTCGAG---GCCATGCCAGCGCGGAGCGGCTTCAGCGCCCCGATG 3218  
 QY 706 ThrLeuThrValAlaValLysLysGluAsnSerPro-----ProArg----- 720  
 Db 3219 ACGCGCAGGTGACGCTGGCGCAGCCCAACAGGGCGCCAGGCCCTTCGCGGCTGGTG 3278  
 QY 721 -----AlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeu 738  
 Db 3279 GGTGACGCTCGGTGGGGCGCGAG-----GTGACGCTC 3311  
 QY 739 AspGlySerArgSerThrAsp---AspGlnArgIleValSerTyrLeuTriPleArgAsp 757  
 Db 3312 GACGCTCGCGTCCAGCGACCGCGATGGCGAGCCCTCACCTTCGGGTGGAGCAGGTG 3371  
 QY 758 GlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGln 777  
 Db 3372 GCGGACCGTCCGCTCACCTCTCCGCTCCGAGGGTGGC-----CAGGTGACGTTCCGC 3425  
 QY 778 LeuThrAsnLeuValGluGlyVal---TyrThrPheHisLeuArgValThrAspSerGln 796  
 Db 3426 GCGCCGAGGTGAGCAGCGCGTGCAGCTTCGAGCTCACCGCGACGCGCGCGC 3485  
 QY 797 GlyAlaSerAspThrAspThrAlaThrValGluValGlnPro 810  
 Db 3486 GCGCGCGCAGCAGCGCGTGTCTCGGGTGACGGTGCAGCCG 3527

RESULT 4

US-09-902-540-2590  
 ; Sequence 2590, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; CURRENT FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 2590  
 ; LENGTH: 4287  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-09-902-540-2590

Alignment Scores:  
 Pred. No.: 1,38e-18 Length: 4287  
 Score: 317.00 Matches: 162  
 Percent Similarity: 38.50% Conservative: 84  
 Best Local Similarity: 25.35% Mismatches: 232  
 Query Match: 5.68% Indels: 161  
 DB: 4 Gaps: 28

US-10-764-390-3 (1-1072) x US-09-902-540-2590 (1-4287)

QY 184 AspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGlyAspSerPro 203

Db 2530 GACCGTCCATCGCGCGGGTGAACCTGTGAGCTCTACGGTTCCGCTGCCGATCCGGAA 2589  
 QY 204 AlaValPro-----AlaGluThrGlnGlnAspProGluLeuHisTyrIleu 218  
 Db 2590 GCGCAGCGGGTGCCTTTCACCTGGGCAAGAGC---TCTGGCCCTGCGGTCACTCGCG 2646  
 QY 219 AsnGluSerAlaSerThrPro-----AlaProLysLeuProGluArgSerValLeu 235  
 Db 2647 GGTCCCAACCCCTGAACCCCGTTCACGGCGCGCAGGTACCGAGTCCACCGCATGT 2706  
 QY 235 ----- 235  
 Db 2707 GTGTTTCAGCTGACCGTCTCCGACGGGTCAAGAGCTCCACGGACTCCGCTGACCGTGACG 2766  
 QY 236 LeuProLeuProThrThrProSerSer-----GlyGluValLeuGluLys 250  
 Db 2767 GTGGCGCTCCCAACAAATCCGCCACCGTGAACCGGGCCTCGACGGCATCGTCCGAGGAG 2826  
 QY 251 GluLysAlaSerGlnLeuGlnGlnSerSerAsnSer----- 263  
 Db 2827 CGCGTGAGTACACGCTGAGCGGCTCCGCCAGCATGCGGATGGCAATGCCCTCACGTAC 2886  
 QY 264 -----SerGlyLysGluValLeuMetProSerHisSerLeuProProLa 278  
 Db 2887 CTGTGGACCCAGGTCTCCGTTACCGCGTGGCGGTGAAGGACTACACCCACCGACGGCG 2946  
 QY 279 SerLeuGluLeuSerSerValThrValGluLysSerProVal-----LeuThrValThr 296  
 Db 2947 ACGTTCATCGCGCGGAGGTACGCTGGATGAGTGTGCTTCCGCTGACGGTCAAGC 3006  
 QY 297 ProGly-Ser-----ThrGluHisSerIleProThrProThrSerAlaAl 312  
 Db 3007 GACGCGATCGCCACGCTGACACGCTCACCCTGACGCTCACCAACCGCAGACCGCGCG 3066  
 QY 312 aProSerGluSerThr-----ProSerGluLeuProIleSerProThrThrAlaPr 329  
 Db 3067 CCCATCGTCTCCGACACGCTCGCTTCGCGCGGCGACGTGTCCACCTGACGGCGCTCC 3126  
 QY 329 aArgThrValLysGluLeuThr-----ValSerAlaGlyAspAsnLeuIleIle--Thr 346  
 Db 3127 GCGGTGCGACCGGATGGTGACGCTCAGCTGCGGAGCAGACCGGTTGTCGACG 3186  
 QY 347 LeuProAspAsnGluValGluLeuLysAla-----PheValAlaProAlaProVal 364  
 Db 3187 GTCCCATCAATGGCGCGCAGCAGTCCGCCATCTCTTCGCCACCCCGTCCCG 3240  
 QY 365 GluThrThrTyrAsnTyrGluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGlu 384  
 Db 3240 ----- 3240  
 QY 385 IleLysGlnGlyHisLysLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrVal 404  
 Db 3241 -----GGTTCGACGAG 3252  
 QY 405 PheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrVal 424  
 Db 3253 TTCACCGTGACGGGACGCGGCTCGGCTCTGCTCCAGGCCGGTG----- 3300  
 QY 425 LysProAlaArgArgValAsnLeuProProValAlaValValSerProGlnLeuGlnGlu 444  
 Db 3301 -----CCGGTGACCATCATCGATGTTCTCTCGCGGGG 3333  
 QY 445 LeuThrLeuProLeuThrSerAlaLeuIleAsp----- 455  
 Db 3334 AACTCTGCGCGACCGCTGACGCGGGCATCGATGCCAGCGGAAACGCTGGTGACGTGCG 3393  
 QY 456 -----GlySerGlnSerThrAspThrGluIleValSerTyrHisTyrGluGlu 472  
 Db 3394 ACGCTCAGCGGCTCCGCGAGCGACCGCGAGGTGACGCTGACGTACCATCTGGGAGCAG 3453  
 QY 473 IleAsnGlyProPheIleGluLysThrSerValAspSerProValLeuArgLeuSer 492

Db 3454 ATCGGTGGC-----ACGGATGTGACCTGACGGCGCGGACACCTGACGCCGTGC 3504  
Qy 493 AsnLeuAepProGlyAsnTySer-----PheArgLeuThrValThrAsp 507  
Db 3505 TTCACCGCGCGGACGAGCGCGGACCGGTACACCGCTCGGCTTTCATCTGACGGTCAGCGAT 3564  
Qy 508 SerAspGlyAlaThrAsnSerThrAlaAlaLeuLeuVal-----AsnAsn 523  
Db 3565 GGCACCTCGACACGAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3624  
Qy 524 AlaValAspTyProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGln 543  
Db 3625 CCGAGCAACACTGCTCCGAGACGCGCGGCGAGTCG-----GCCATCTGCGCGGAGGC 3681  
Qy 544 AsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuTyrglu 563  
Db 3682 GCCACTGTGAGCTCAACGCGACGCGGACCGAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3741  
Qy 564 TrpSer---LeuGlyProGlySerGluGlyHisValValMetGlnGlyValGln--- 581  
Db 3742 TGGACGCGAGATCGGC-----GGCAGCGCGGTGAGCGGTGCGGTGCGGTGCGGTGCGGT 3789  
Qy 582 ---ThrProTyLeuHisLeuSerAlaMetGlnGlyAspTyThrPheGlnLeuLys 600  
Db 3790 CTGACGCGCGAGTTCACCGCGCGCGCTCTCCCGAT---CCGCTGACCTTCTCTCTGATG 3846  
Qy 601 ValThrAspSerArgGlnGlnSerThrAla---ValThrValIleValGlnProGlu 620  
Db 3847 GTACGCGACGCTAGC-----GCGACTGCTGCGACGTGAGCTGCTGCTGCTGCTGCTGCTG 3900  
Qy 621 AsnAsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGlu 640  
Db 3901 GAGAACGTCGCGCGGTGCGGTGCGCT-----CGCGCGGTGCTCTCTGCGCAACGACGACG 3954  
Qy 641 SerAlaThrLeuAspGlySerSerSerAspAspHisGly---IleValPheTyHis 659  
Db 3955 TCGCGCCAGCTCGACGCTCGGCTTCAGCGACGCGGACGCGGTGCTGCTGCTGCTGCTGCTG 4014  
Qy 660 TrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAla 679  
Db 4015 TGGACGCGAGGTCTCGCGCGCGGACGCGAC---ATCTCCGCGCGGACGCGGTGCTGCTG 4071  
Qy 680 ThrValThr-----GlyLeuGlnValGlyThrTyHisPheArgLeuThrVal 695  
Db 4072 GTGCTCAACCTGCTGACCTCGACGACGACGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 4131  
Qy 696 LysAspGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysLysGluAsn 715  
Db 4132 AAGGACTCCGCGGACCGAGAGGACGACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4191  
Qy 716 AsnSer-----ProProArgAlaArgAla-GlyGlyArgHis 727  
Db 4192 GACAGCGGTGCTGCTCGCGACGCGCGGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4238

## RESULT 5

US-09-902-540-1119  
; Sequence 1119, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1119  
; LENGTH: 16584  
; TYPE: DNA

; ORGANISM: Myxococcus xanthus  
US-09-902-540-1119  
Alignment Scores:  
Pred. No.: 1,56e-17 Length: 16584  
Score: 317.00 Matches: 162  
Percent Similarity: 38.50% Conservatve: 84  
Best Local Similarity: 25.35% Mismatches: 232  
Query Match: 5.68% Indels: 161  
DB: 4 Gaps: 28  
US-10-764-390-3 (1-1072) x US-09-902-540-1119 (1-16584)  
Qy 184 AspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGlyAspSerPro 203  
Db 6679 GACCGGTTCATCGCGCGCGGTGAACCTGGTGAAGCTCTACCGGTTCGGCTGCCGATCCGGAA 6738  
Qy 204 AlaValPro-----AlaGluThrGlnGlnAspProGluLeuHisTyLeu 218  
Db 6739 GGCAGCGCGGTGACCTTCACCTGGGCACAGAG---TCTGGCCTTGGCGTCACCTCGCG 6795  
Qy 219 AsnGluSerAlaSerThrPro-----AlaProLysLeuProGluArgSerValLeu 235  
Db 6796 GGTCCCAACACCTGAACCCGTCGTTACGCGCGCGGAGGTTACCGAGTCCACCGCACTG 6855  
Qy 235 ----- 235  
Db 6856 GTGTTACGCTGACCGTCTCCGACGGGTCAAGAGCTCCACGGACTCCGCTGACCGTGACG 6915  
Qy 236 LeuProLeuProThrThrProSerSer-----GlyGluValLeuGluLys 250  
Db 6916 GTGCGGTTCCTCCCAACATCCGCCCGGTGACCGCGGCTGACCGGCTGCGGATCGTCGAGGAG 6975  
Qy 251 GluLysAlaSerGlnLeuGlnGlnSerSerAsnSer----- 263  
Db 6976 CGCGCTGAGTACACGCTGAGCGGTTCGCCAGCATGCGGATGCGATGCGCTCCTCAGCTAC 7035  
Qy 264 -----SerGlyLysGluValLeuMetProSerHisSerLeuProProAla 278  
Db 7036 CTGTGGAGCCAGGTCTCCGCTACCGGTGCGGTGAGGACTACACCGCGCGCGCG 7095  
Qy 279 SerLeuGluLeuSerSerValThrValGluLysSerProVal-----LeuThrValThr 296  
Db 7096 ACGTTCATTCGCGCGCGGAGTACCGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7155  
Qy 297 ProGly-Ser-----ThrGluHisSerIleProThrProProThrSerAlaAl 312  
Db 7156 GACGCGATCGCCACGCTGACGACACGCTGACCGGTGACCGGTGACCGCGCGCGCG 7215  
Qy 312 aProSerGluSerThr-----ProSerGluLeuProLysSerProThrThrAlaPr 329  
Db 7216 CCATCGTCTCCGACACGCTGCGGTTCGCCGCGGCACTGTACCGTACCGGCTCC 7275  
Qy 329 aArgThrValLysGluLeuThr-----ValSerAlaGlyAspAsnLeuIleIle--Thr 346  
Db 7276 GCGGTGACCGCGGATGTCAGTCCGCTGACGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7335  
Qy 347 LeuProAspAsnGluValGluLeuLysAla-----PheValAlaProAlaProProVal 364  
Db 7336 GTCCCATCATGTCGCGCGGACACCGTCCGCCATCTCTTCGCCACCCCGGTCCCG----- 7389  
Qy 365 GluThrThrTyAsnTyArgGluTrpAsnLeuIleSerHisProThrAspTyGlnGlyGlu 384  
Db 7389 ----- 7389  
Qy 385 IleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyVal 404  
Db 7390 -----GGTTCGACGAG 7401  
Qy 405 PheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrVal 424  
Db 7402 TTCACGTGACGCGGACGCGGCTGCGCTCTGCTCTCAAGCGCGGTG----- 7449

QY 425 LysProAlaArgValAsnLeuProProValAlaValValSerProGlnLeuGlnGlu 444  
 Db 7450 -----CGGTGACCATCATCATGTTCTTCGCGCGG 7482  
 QY 445 LeuThrLeuProLeuThrSerAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 455  
 Db 7483 AACTCTGCGCGCGCGCGCATCGATGCGCGCGCGCGCGCGCGCGCGCGCG 7542  
 QY 456 -----GlySerGlnSerThrAspThrGluLeuValSerThrHisTrpGluGlu 472  
 Db 7543 ACGCTCAGCGGCTCCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7602  
 QY 473 IleAsnGlyProPheIleGluLysThrSerValAspSerProValLeuArgLeuSer 492  
 Db 7603 ATCGGTGGC-----ACGATGTGACCTTCAGCGGCGCGCGCGCGCGCGCG 7653  
 QY 493 AsnLeuAspProGlyAsnTyrSer-----PheArgLeuThrValThrAsp 507  
 Db 7654 TTCAACCG 7713  
 QY 508 SerAspGlyAlaThrAsnSerThrThrAlaAlaLeuLeuLeuLeuLeuLeuLeu 523  
 Db 7714 GGCACCTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7773  
 QY 524 AlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGln 543  
 Db 7774 CCGAGCAACACTGCTCCGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7830  
 QY 544 AsnSerIleThrLeuAsnGlyAsnGlnSerSerAspHisGlnIleValLeuTyrGlu 563  
 Db 7831 GCCACTGTGACGCTCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7890  
 QY 564 TrpSer---LeuGlyProGlySerGluLysHisValValMetGlnGlyValGln--- 581  
 Db 7891 TGGACGCGATCGC-----GGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 7938  
 QY 582 ---ThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLys 600  
 Db 7939 CTGACGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7995  
 QY 601 ValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnProGlu 620  
 Db 7996 GTCAGCGCGCGGTACG-----GGACTGTGTCGACGTGACGTCCATCATCGCGCG 8049  
 QY 621 AsnAsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGlu 640  
 Db 8050 GAGAACGTCGCGCGCGGTGCGCGCT-----CGCGCGGTGCTCTCTGGCAACCGACG 8103  
 QY 641 SerAlaThrLeuAspGlySerSerSerAspAspHisGly---IleValPheTyrHis 659  
 Db 8104 TCGGCCACGCTCGACCGGCTCGGCTTCACGCGCGCGCGCGCGCGCGCGCGCG 8163  
 QY 660 TrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAla 679  
 Db 8164 TGGACGCGGTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8220  
 QY 680 ThrValThr-----GlyLeuGlnValGlyThrThrHisPheArgLeuThrVal 695  
 Db 8221 GTGCTCAACCTGCTGACCTCGACGACGACGCGCGCGCGCGCGCGCGCGCG 8280  
 QY 696 LysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysLysGluAsn 715  
 Db 8281 AAGGACTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8340  
 QY 716 AsnSer-----ProProAlaAlaArgAla-GlyGlyArgHis 727  
 Db 8341 GACAGCGGTGGCTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8387

RESULT 6

US-09-902-540-6552  
 ; Sequence 6552, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiesand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; CURRENT FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 6552  
 ; LENGTH: 3282  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 ; US-09-902-540-6552

Alignment Scores:  
 Pred. No.: 1,59e-14 Length: 3282  
 Score: 271.50 Matches: 207  
 Percent Similarity: 34.89% Conservative: 107  
 Best Local Similarity: 23.00% Mismatches: 304  
 Query Match: 4.87% Indels: 284  
 DB: 4 Gaps: 49

US-10-764-390-3 (1-1072) x US-09-902-540-6552 (1-3282)

QY 16 ThrIleAlaGlyCysAlaArgLysGlnCysSerGluGlyArgThrTyrSerAsnAlaVal 35  
 Db 842 ACTACTACGGGTGCGCGCGCGCTTACGCCACCGCGCGCGCGCGCGCGCGCG 901  
 QY 36 IleSer-----ProAsnLeuGluThrThrArg--- 44  
 Db 902 ACGCGCTGACCTTCGGGCACATTCAGAACGCGCAATCCCTCCGACGAGCCACCGCTCG 961  
 QY 45 -----IleMetArgValSerHisThrPheProValValAspCysThrAlaAlaCys--- 61  
 Db 962 CCTACGTCAGCGCGCGCTCGGCAACTCCGCGG-----TGCACAGCAGCGCGGAGG 1012  
 QY 62 CysAspLeuSerSerCysAspLeuAlaTrpTrpPheGluGlyArgCysTyr----- 78  
 Db 1013 TGTGG-GCCTCC-----ATGCTCTGG-----GAGTGTACGTGTGCTC 1050  
 QY 79 LeuValSerCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSer 98  
 Db 1051 CTCAGGCGCCACCCCTTCAGAGGCCCGAGCCGATGACACTACTGCTGTCGCGCG 1110  
 QY 99 TyrLeuThrPheValLeuArgPro----- 106  
 Db 1111 TACAGGCCACCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1170  
 QY 107 -----ValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgGly 124  
 Db 1171 ACCCGCGCGAGCGACCGCGCGCGCTTCTCCAGCGCTTCCTCCAGCGCTTCGCCAAGCGCG 1230  
 QY 125 SerProSerGlyIleTrpGly-----AspSerProGluAspIleArgLysAspLeu 141  
 Db 1231 GCGCGCATGGGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1290  
 QY 142 \*\*\*PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTyr\*\*\*AspAspTyrArg 161  
 Db 1291 AGCTTTCGCGCGCGCGCAATTC---ATCAGAGGTACCTCCCTCGCGGTTCGATCAGCGCGCT 1347  
 QY 162 Glu----LeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAla 180  
 Db 1348 GAGGCGTCGCAAGGAC----- 1374  
 QY 181 GluTyrThrAspTrpGlyLeuLeu-----ProGlySerGlu 192  
 Db 1375 GACGTGGCGGAGACCGGCTGCTGCGGTGACGTGCGCAACGTGGGCGCGGTTGCTGCTG 1434  
 QY 193 GlyAlaPheAsn----- 196

Db 1435 GCGCCTTCTCCGGCAGCGTCTCCACCACAGCAGCAGCGCCAGCTGGAGTTCCCTCG 1494  
Qy 197 -----SerSerValGly----- 200  
Db 1495 GGCACACCCCTGACCTTACGTCGATGCGCGCGCGCAGCGCCAGCGCCACCTCCCG 1554  
Qy 201 -----AspSer 202  
Db 1555 GTGAGTTGGCGCGCGCGCAGCCGACGACACGCGCGCTGCGCATCACCTTCGACGAG 1614  
Qy 203 ProAlaValProAla-----GluThrGlnGlnAspProGluLeuHisTyr----- 217  
Db 1615 CCCTCGCTCCCGCGCAGCGCAGCAGCAGCTTTCAGCCCTCGGCTGAAGTACGACGAA 1674  
Qy 218 LeuAsnGluSerAlaSer-ThrProAlaProLysLeuProGluArgSerValLeuLeuPr 237  
Db 1675 GTGCCCGCGCGCGCGGAGGACGACTTCGAGGCGCGCTGTGTCATGACCTCGCTCC 1734  
Qy 237 oLeuProThrThrProSer-----SerGlyG1 246  
Db 1735 ACCTGGCTCAACCCCTACGCGGAGCGGACTGGAAGGTCAGGGTCCGAGGGCGGCGC 1794  
Qy 246 uValLeuGluLysGluLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLY 266  
Db 1795 TACATGACCGCCCAACCCCGCAGTACAGCGGACCTCTCTCATCTCCCTCGATG 1854  
Qy 266 sGluValLeuMetProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValTh 286  
Db 1855 AAGGTGAATGAGACCGCG-----CCTTCTGTTTCAGCTTCGCTACCGC 1899  
Qy 286 rValGluLysSerProValLeuThrValThrProGlySerThrGluHisSerIleProTh 306  
Db 1900 CACTCGTTCGAGTCCG-----ACATCTACCCCGCGCGCT-----CCATC 1938  
Qy 306 rProProThrSerAlaAlaProSerGluSerThrProSerGluLeuProIleSerProTh 326  
Db 1939 TTCCCGTACTACGCGCGCGCTCTCGAGTTACCCACG---ACGCGCTCGAGTGTAC 1995  
Qy 326 rThrAlaPro-----ArgThrValLysG1 334  
Db 1996 GACATCGCGCGATGGCGTGGCTCCATCTACTCCATCGACCGCGAGCGGTCCCG 2055  
Qy 334 uLeuThrValSerAlaGlyAspAsnLeuIleThrLeuProAspAsnGluValGluLe 354  
Db 2056 GGCTACATCGAGCAGGGCG----- 2074  
Qy 354 uLysAlaPheValAlaProAlaPro-ProValGluThrThrTyrAsnTyrGluTrpAsnL 374  
Db 2075 -----CCAACCCGCTGATGTACCGCAACGC----- 2099  
Qy 374 euLleSerHisProThrAspTyrGlnGlyGluLysGlnGly----- 388  
Db 2100 --GTGGGTGGCCAGAACGTACGACGAGAGAACGCAAGCGGTGCGCTTCCCGAGTGG 2157  
Qy 389 --HisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrVal---PheLysV 407  
Db 2158 TCCACCGCGCAG--CTGGCGTGGCGACATGTTCCGCGGCAAGACGTCCAGTTCGCT 2215  
Qy 407 alThrValSerSerGluAsnAlaPheGlyGlu-----GlyP 419  
Db 2216 TCCGATTTGTACCGACACCGCGTGGCGGTGCGGTTCGAGCTTCGACGACGAGTGGCT 2275  
Qy 419 heValAsnValThrValLysPro---AlaArgArgVal----- 430  
Db 2276 TCACCAAGTGGAGCGCATGCTTCAGCGCGCGGTGGACGAGCGGCGACGCGGCCA 2335  
Qy 431 -----AsnLeuProProValAlaVal-----SerProG 441  
Db 2336 CCTGCAACCGCGCGGTGGCCAAACCGCGCGCGACGTCCGCGACGCCCGCGCG 2395  
Qy 441 lnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeu-----IleA 455  
Db 2396 TGTATGACGCGGTGAGAAACCCGAGACGCGGGGCACTGGAGCTCGTGGCCACACGCTGG 2455

Qy 455 spGlySerGlnSerThrAsp---AspThrGluLysValSerTyrHisTrpGluLysLeA 474  
Db 2456 ACGCACCGCGCAGCTATGACCCGAGGCGCCAGCCCTCACCCTACCTACCTGGACGAGTGA 2515  
Qy 474 enGlyProPheIleGluLysThrSerValAspSerProValLeuArgLeuSerAsnL 494  
Db 2516 GCGGG-----CCGCGGTACCTTGGTGAACG 2542  
Qy 494 euAspProGlyAsnTyrSerPheArgLeuThrValThrAspSer----- 508  
Db 2543 CGGACACCGCCACCCCGTCTGTTACGCGCGATGTGCCGACGACTCCATCTCCTGACGTTCC 2602  
Qy 509 -----AspGlyAlaThrAsnSerThr-----ThrAlaLysLeuLys 521  
Db 2603 AGCTCGTGGTGGCGATGGCTCGAGACCGCGCGGACACACGAGTCTCGGCGCTGGGTCC 2662  
Qy 521 alAsnAsnAlaValAspTyrProProValAlaLysGlnGlyProAsnHisThrIleThrL 541  
Db 2663 TCACAC-----GTCACACCGCGCCCCACGCGGCTGGTGACCG 2698  
Qy 541 euPro-----GlnAsnSerIleThrLeuAsnGlyAsnGlnSerS 554  
Db 2699 CGCGCGGAGGTGGCGAGTTCCTCTCGGAGCCGCTGACGCTGACGCCCGCGCTCGA 2758  
Qy 554 erAsp---AspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyL 573  
Db 2759 GCGACCGCGAGCGGAGCTCATCTCTACAAAGTGGCG-----CAGACGTCCTCGGTC 2809  
Qy 573 yHisValValMetGln-----GlyValGlnThrProTyrLeuHisLeuSerAlaMetG 591  
Db 2810 CCGCGCTCGCGCTCAAGCGGGCTCCAGCGCCACGCGAGTTCAGGTGCCGAGGTGG 2869  
Qy 591 ln---GluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerT 610  
Db 2870 CCGTCGATACTCACTTCACTTCACTGCTGGTGGTGAACCACTGCGAGGACCCAGAGCAAGA 2929  
Qy 610 hrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaG 630  
Db 2930 CGGCACCGCTGACGTGATGTGGCG---AACGTGACCGGCGCCCGTGGCGATGCGG 2986  
Qy 630 lProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerS 650  
Db 2987 GC---TCCGACAGTGGTGACCCGCTGGCCACCGGAGCTCAGCGCAGCGCTGTGTG 3043  
Qy 650 erAspAspHisGlyIleValPheTyrHis-TripGluHisValArgGlyProSerAlaVal 669  
Db 3044 GACCCGACGCGGATGCCCTCACCCTGACGTGGCGTCACTGGAGGGGCCCC---GCGGTG 3100  
Qy 670 GluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnVal----- 686  
Db 3101 GAGTCGAGGAGCGGACAGCGCCACCGCTGCTTCAACCGCCCGAGGTGGCGCGGAG 3160  
Qy 687 GlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThr---Ser 705  
Db 3161 ACGETGTGCTGCTTCGAG---GCCATCGCCAGCGCAACCGGCTGTCCAGCGCCCCCGATG 3217  
Qy 706 ThrLeuThrValAlaValLysLysGluAsnSerPro-----ProArg 720  
Db 3218 ACGCGAGGTGACGTTGGCGCAGCCCAACAGGGCGCCCGGCTTCCCGGG 3271

## RESULT 7

US-09-248-796A-3204  
; Sequence 3204, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: US 60/096,409  
 ; PRIOR FILING DATE: 1998-08-13  
 ; NUMBER OF SEQ ID NOS: 28208  
 ; SEQ ID NO 3204  
 ; LENGTH: 2481  
 ; TYPE: DNA  
 ; ORGANISM: Candida albicans  
 US-09-248-796A-3204

## Alignment Scores:

Pred. No.: 1.73e-09 Length: 2481  
 Score: 215.50 Matches: 198  
 Percent Similarity: 32.80% Conservative: 91  
 Best Local Similarity: 22.47% Mismatches: 355  
 Query Match: 3.86% Indels: 239  
 DB: 4 Gaps: 40

US-10-764-390-3 (1-1072) x US-09-248-796A-3204 (1-2481)

QY 11 LeuLeuLeuVal-ThrIleAlaGlyCysAlaArgLysGlnCysSerGluGlyArgTh 30  
 DB 260 CTACTTCTACTCACCACAGCACACCGGGTGAACATCAAAACATGTGAACACAAAATCAT 319  
 QY 30 rTyrSerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSerHi 50  
 DB 320 CTCTCAGATCAATC-----ACCTCCACATCTCCCA 352  
 QY 50 sThrPheProValAlaPheCysThrAlaAlaCysCysAspLeuSerSerCysAspLeuAl 70  
 DB 353 CTCTCTAC-----TGTTGCTGCTGCTGCAC----- 380  
 QY 70 atrPhePheGluGlyArgCysTyrLeuValSerCysProHisLysGluAsnCysGluPr 90  
 DB 381 ----TTCTTCAAGTCCTGTGCTAGTACTCGCTCCTGTT-CAACTTCTGAACAAAGCAA 435  
 QY 90 oluLysMetGlyProIleArgSerTyrLeuThrPheValLeuArgProValGlnArgPr 110  
 DB 436 CAAGAGAGAGTACAGCTAGACATCAACTT---CACAGCTACACAGCTACTACTTCC 492  
 QY 110 oAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgGlySerProSerGlyIleTr 130  
 DB 493 AATACTCTCTCTT----- 505  
 QY 130 pGlyAspSerProGluAspIleArgLysAspLeu\*\*\*PheLeuGlyLysAspTrpGlyLe 150  
 DB 505 ----- 505  
 QY 150 uGluGluMetSerGluTyr\*\*\*AspAspTyrArgGluLeuGluLysAspLeuGlnPr 170  
 DB 506 -----CTCTTCCACC 516  
 QY 170 oSerGlyLysGlnGluProArgGlySerAlaGluTyrThrAspTrpGlyLeuLeuProGl 190  
 DB 517 TCCAGGAAACCCCACTCAAC-ACCTGCTCAG----- 549  
 QY 190 ySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThrGl 210  
 DB 550 -ACTAGTTCTGCCAACAAACAATCAATCATCAATACAGAGCTCCATCTACTAGTGT 608  
 QY 210 nGlnAspPro-----GluLeuHisTyrLeuAsnGluSerAlaSer----- 223  
 DB 609 TATCCAAACCAAGTACATCAGAGTTCTATGTTCAATCACACAAACTTCTACACACCAA 668  
 QY 224 -----ThrProAlaProLysLeuProGluArgSerValLeuLeuPro----- 237  
 DB 669 CACTCCAACACTAGTCTACCAATACTCCAACTTCCAGAGCAGCAGCACTACTTTCAGC 728  
 QY 238 -----LeuProThrThrProSerSerGlyGluValLeuGluLysGl 251  
 DB 729 AGCACCACACATCAGAGGCTCCAGTTACTCCAGTAGTACCAAGTAGTACCAANATAC 788  
 QY 251 ulysAlaSerGlnGlnGlnSerSerAsnSerSerGlyLysGluValLeuMetPr 271

DB 789 ACCAACCCACA-----TCAGAGGACCAAAATACA-----CC 818  
 QY 271 oSerHisSerLeuProAlaSerLeuLeuSerSerValThrValGluLysSerPr 291  
 DB 819 AACCATATCAGAGGCTCCAGTTACTCCAGTACTTCAAGATTA---GTACCAAAATACACC 875  
 QY 291 oValLeuThrValThrProGly-----SerThrGluHisSerIleProThrProProTh 309  
 DB 876 AACCATTTCAAAGCACCAAAATACACCAACCACTCAGAGCAGCAGCTACCAACACAC 935  
 QY 309 rSerAlaAlaProSerGluSerThrProSerGluLeuProLysSerProThrAla--- 328  
 DB 936 TTCAGAGCACCAAAACACTCCAAACCACTCAGAGGCTCCGTTACTTCAACACCACTCAGA 995  
 QY 329 -----ProArgThr-----Vally 333  
 DB 996 AGTAGTACCAACCACTTCAACTCAGAGGAGTGTGTATCAACTTCATCACTAGTGTATC 1055  
 QY 333 sGluLeuThrValSerAlaGlyAspAsnLeuIle-----IleThrLeuProAspAs 350  
 DB 1056 CGAACAGACAACTTTAACTTCTACTTCTTACTTCACTCTGCTCTGCTACACACA 1115  
 QY 350 nGluValGluLeuLysAlaPheValAlaProAlaPro-----ProValGluThrTh 367  
 DB 1116 GACTAGTACTCCAGAGGCTTCGATTTCTCTAAACCAAGTTCAACTTCTATTGAAACACC 1175  
 QY 367 rTyrAsnTyrGluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluLeuLysGl 387  
 DB 1176 ATCAACTTCAACTTTTGAACAGATCCAACTACTTCCAGTGTGGTACCCCTTCATC 1235  
 QY 387 nGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysVa 407  
 DB 1236 GGAGCAACTCAACCAACTACTTCTTGAAGTGGCT-----GT 1274  
 QY 407 lThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAl 427  
 DB 1275 AACATCAATTTCTCAACTCAAGATCAACTTCTTGGTTGAACCAACCAACTAGTCTTT 1334  
 QY 427 aArgArgValAsnLeuProValAlaValValSerProGlnLeuGlnGluLeuThrLe 447  
 DB 1335 GGAAAGTTCCCAACACTCCA-----ACACCAACCCCTTCCACTTCTGAA---GCCCA 1382  
 QY 447 uProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSe 467  
 DB 1383 ACCATCACTAGTGA-----AGTCAAGCTCCACAGATACC----- 1419  
 QY 467 rTyrHisTrpGluGluIleAsnGlyProPheIleGluGluLysThrSerValAspSerPr 487  
 DB 1420 -----ACTTCTAGTGCACAGCACC 1439  
 QY 487 oValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArg----- 502  
 DB 1440 AGAAGTTCTTCAAGTAAATGCAGATTTCTCCAACTCGTATTGTCATAGTTCGGAACACC 1499  
 QY 503 -----LeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuI 520  
 DB 1500 TTCATTAGTGAACCTTACCGATTCCAAATTTGATAGTCTTCCACTACTGATGCC----- 1554  
 QY 520 eValAsnAsnAlaValAspTyrProProValAlaAlaAsnAlaGlyProAsnHisThrIleTh 540  
 DB 1555 -GTTTTCGCAAGCTACTACTGAAACCGACTTCTGAAACACACA---CCTACTGTGCTTCTTC 1610  
 QY 540 rLeuProGlnAsnSerIleThrLeu-----AsnGlyAsnGlnSerSe 554  
 DB 1611 TGTAACTGCCAATGACATTAATTCAGCCCAATATTTCAGCTCCAACTTCTTAATGCTGACGC 1670  
 QY 554 rAspAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyLysHi 574  
 DB 1671 CGAAACTGCTTCACTCCAGTTTCTGAAACAAAGTTTAGCCACTGGAAGTCAACACAGTCT 1730  
 QY 574 sValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluLysAs 594



```

Db 1731 TGATACTACTGCTGGCGCTAGTCTCAACTGCCAGTGAAGCAACTGCT-----GAGAAATT 1784
Qy 594 pTyThrPheGlnLeuValThrAspSerSerArgGlnGlnSerThrAla***ValTh 614
Db 1785 GTCTACTTTTGGT-----ACTGATGGCTCTTCAGATGCTTCT-----CA 1823
Qy 614 rValIleValGlnProGluAsnAsnArgPro-----ProValAlaVa 628
Db 1824 AACATTGCCGAACACCAGCAATTTCTCTGATCAATCAGTGGTTACACCTCAGCTAG 1883
Qy 628 lAlaGlyProAspLysGluLeuIlePheProVal-----GluSerAlaThr-----Le 644
Db 1884 TGCTTCTCCAGATGATATACACTC---CCAACTGGATCTGAAAGTGTGACACGCTTGT 1940
Qy 644 uAspGlySerSerSerSerSerAspHisGlyIleValPheTyThrHisTrpGluHisValAr 664
Db 1941 TTCTGGATCTGAACACGATATTGATCAACATACAGTT-----GCTTC 1982
Qy 664 gGlyProSerAlaVal---GluMetGluAsnIle-----674
Db 1983 AGGATCTACTGTTATTCAGAAAGCAGTAATTTCCACCCCAATCACCATCACAATCTGT 2042
Qy 675 -----AspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyThrHisPh 691
Db 2043 AGTATCATCAGATGCTGCTCTTAACGTCTCAACAGTTTCAGAACCACTGATTCTCT 2102
Qy 691 eArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThr-----Va 709
Db 2103 TGCTGGAAGTGAACACCGAGTACACCAATATCATCAAGTCTACAGTACTAGTGAACC 2162
Qy 709 lAlaValLysGluAsnAsnSerProArgAlaArgAlaGlyGlyArgHisValLe 729
Db 2163 AGTATTTTCAAGTCAGTATTAATTC-----AGTGAAGTACCACTAGTTT 2207
Qy 729 uValLeuProAsnAsn-----SerIleThrLeuAspGlySerArgSerThrAspAs 746
Db 2208 GGTGTTCTTACCATTCTGAACCTAGCTCAACTGTTACTGTTAGTTCTGAACTGTGTC 2267
Qy 746 pGlnArgIleValSerTyLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspVa 766
Db 2268 TACTGCTATTAAATCTGAATCGGTCTTAACGTGTTCTGATAGTCTGCT-----AC 2321
Qy 766 lIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTy 786
Db 2322 AGTTACTGGTAGTGAATCAATTTTGACTGGTGAATPACTGAAACAAGTGTCTACTGTATGTC 2381
Qy 786 rThrPheHisLeuArgValThrAspSerGlnGlyValaSerAspThrAspThrAlaThrVa 806
Db 2382 AAGTGAATCTACTTTAACTGATCCACTACTGTTGCTACTGCTCCGCTGCACACTACTAT 2441
Qy 806 l 806
Db 2442 T 2442

```

RESULT 8

```

US-09-902-540-2872
; Sequence 2872, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2872
; LENGTH: 1119
; TYPE: DNA

```

```

; ORGANISM: Myxococcus xanthus
US-09-902-540-2872
Alignment Scores:
Pred. No.: 2,34e-09 Length: 1119
Score: 207.50 Matches: 95
Percent Similarity: 41.40% Conservative: 59
Best Local Similarity: 25.54% Mismatches: 158
Query Match: 3.72% Indels: 60
DB: 4 Gaps: 17

```

US-10-764-390-3 (1-1072) x US-09-902-540-2872 (1-1119)

```

Qy 410 SerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgArg 429
Db 52 ACGGATGAGCGCGCGGGGCTTTGGCTGTTTCATCGATGACATCGCTTCACGGGGTTG 111
Qy 430 ValAsnLeuProProValAlaValSerProGln-----441
Db 112 CGNACACGCCCTTCACGGCGGTGTGACGACGAGGGGTCTGCACGACGGGCATGTG 171
Qy 442 -----LeuGlnGluLeuThrLeuProLeuThrSerAla 452
Db 172 GATGCGCAGCGCGCGAGGACCTGGCGGTTCGATGAGCTGTCCGTG-----GTG 219
Qy 453 LeuIleAspGlySerGlnSerThrAspAspThrGluIleValSerTyThrTrpGluGlu 472
Db 220 CAACTGACGGCGACGCGCAGCGCTGGCGGTGTGCTCAGCTACCGCTGGGAGCAG 279
Qy 473 IleAsnGlyProPheIle-----GluGluLysThrSerValAspSer 486
Db 280 GTGCGGGTCCCTTCGCGCGATGATGGCGCGACAGCGCCAGCCGCTTCCACGCG 339
Qy 487 ProValLeuArgLeuSerAsnLeuAspProGlyAsnTyThrSerPheArgLeuThrValThr 506
Db 340 CCGAGGTGTGCGCGCATACATG-----CTGACCTTCGCTGACGGTGTG 387
Qy 507 AspSerAspGlyAlaThrAsnSerThrAlaAlaLeuIleValAsnAsnAlaValAsp 526
Db 388 GACGGC---GGTCTGCGGACAGCGACAGCTGCAGGTGTCTGTCGCCAG---GTGAAC 441
Qy 527 TyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIle 546
Db 442 AAGCGCGCGGGTGTCCGCGGACCCGCGCAGACGGTGAC---GAGGGGAGCAGCGGTG 498
Qy 547 ThrLeuAsnGlyAsn---GlnSerSerAspHisGlnIleValLeuTyThrTrpSer 565
Db 499 ACGCTCCAGGAGAGCGGAGGACCGCGACGGCGCGGTGTGTCAGCCAGCAGTGGACG 558
Qy 566 LeuGlyProGlySerGluGlyLysHisValValMetGlnGlyValGln-----ThrPro 583
Db 559 -----CAGGTCTCCGCGCTCCGCTCAGCTGACGGCGCTGATACGCTGACCCCG 609
Qy 584 TyrLeuHisLeuSerAlaMetGlnGluGly-----AspTyThrThrPheLeuLysVal 601
Db 610 AGCTTACCGCGCGCGGTGCGCCAGAGGAGCGGAGCTGCTCTTCCAGCTGTGTGTG 669
Qy 602 ThrAspSerSerArgGlnGlnSerThrAla***ValThrValIleValGlnProGluAsn 621
Db 670 AGTGACGGGCGAGTTGAGTGGCGAGGCGCGCGCGGTGAGTGTGACGGTGGCG---720
Qy 622 AsnArgProProValAla-----ValAlaGlyProAspLysGluLeuIlePheProVal 639
Db 721 ---CACGTGCGCTGCGCGCCACCGTGAAGCGCGGGGACGATGTGGCCACCTTCTCGCG 777
Qy 640 GluSerAlaThrLeuAspGlySerSerSerSerAspHisGlyIleValPheTyThrHis 659
Db 778 AAGTCCGTACGTTGACGGCCACCGGACCGGACCGGAGCGGAGCGAGCTAGCTAGCGG 837
Qy 660 TrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAla 679
Db 838 TGGAGCCAGCGGAGGGGCCCCACC---GTGACGCTCCAGAACGCGACGCGCGCGGTGTC 894

```



; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 8.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/453,702B  
 ; FILING DATE: 03-Dec-1999  
 ; CLASSIFICATION: <Unknown>  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/110,955  
 ; FILING DATE: 04-DEC-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J.  
 ; REGISTRATION NUMBER: 27386  
 ; REFERENCE/DOCKET NUMBER: 960296.95017  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (608) 251-5000  
 ; TELEFAX: (608) 251-9166  
 ; INFORMATION FOR SEQ ID NO: 39:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 25165  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 39:

US-09-453-702B-39  
 Alignment Scores:  
 Pred. No.: 2,27e-06 Length: 25165  
 Score: 201.50 Matches: 157  
 Percent Similarity: 34.40% Conservative: 102  
 Best Local Similarity: 20.85% Mismatches: 283  
 Query Match: 3.61% Indels: 211  
 DB: 3 Gaps: 30

US-10-764-390-3 (1-1072) x US-09-453-702B-39 (1-25165)

QY 190 GlySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThr 209  
 DB 14920 GGATCTGATCTGGCGCTT---TCTGGCACCAGCGATCATCGCTGGCGGATCGCGGATCAC 14976  
 QY 210 GlnGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeu 229  
 DB 14977 -----GTCACTTACCGGACAACTACAGCGCCACCGATGCTCTCC----- 15021  
 QY 230 ProGluArgSerValLeuLeuProLeuProThrProSerSerGlyGluValLeuGlu 249  
 DB 15022 GCAACTGGAGTGTACCGTCTGCTCGGGGTCAGTGGCTGGCGGAGCGACCTAC 15081  
 QY 250 LysGluLysAlaSerGlnGlnGlnGlnSerSerAsnSerSerGlyLysGluValLeu 269  
 DB 15082 AGCGTGACGGCGGCGTCACTCAATGCTCAGGGTAACAGCAGCAGCCAGCCGATCAACGTG 15141  
 QY 270 MetProSerHisSerLeuProAlaSerLeuGluLeuSerSerValThrValGluLys 289  
 DB 15142 CAGGTTAATACCGCTGCGGGGATCACCATTAT-----CCGGTGGCGAGC 15189  
 QY 290 SerProValLeuThrValThrProGlySerThrGluHisSerIleProThrProThr 309  
 DB 15190 GACGATATTATCAACGCTTCAGAGCGCAGCGCGCAACTATCAGCGCCAGGTAAAC 15249  
 QY 310 SerAlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaPro 329  
 DB 15250 GGTGGCGCGCGGCGAGCAGCCGCTCACTGTAAGTGGGGGTAAACTTACACCGCC--- 15306  
 QY 330 ArgThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleThrLeuPro--- 348  
 DB 15307 -----ACCGTCCAGGCTGATTTAAGCTGGAATGTGAGCGTACCTGCC 15348  
 QY 349 -----AspAsnGluValGluLeuLysAlaPheValAla----- 359  
 DB 15349 GCGGACTGCGAGGCTCTGGGTAAACGGTGAATTAACGGTTAATGCTCGGTGACCAACGCC 15408  
 QY 360 -----ProAlaPro 362

DB 15409 GTTGGCAACACCGGTAGCGGACCGCGGATATACCATGATGTCAGTCTGCTGGCGCTG 15468  
 QY 363 ProValGluThrThrTyrAsnTyrGlu---TrpAsnLeuIleSerHisProThrAspTyr 381  
 DB 15469 CGGGTGGATACCGTCCGGGGTGATGATGTGTGTCATATATCATCGAACACGCT- 15519  
 QY 382 GlnGlyGluIleLysGlnGly----- 388  
 DB 15520 CAGGCGCAGGTGATCACCAGCAGCGAGCTCCGGCTTTCGGCGAGCAGCGGCTTACCGGTG 15579  
 QY 389 -----HisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLys 406  
 DB 15580 GTGATTAAACAACCAACTTACGCCGCCACGGTGTGGCGAAGCTAGCTGGAGCGTCCGC 15639  
 QY 407 ValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysPro 426  
 DB 15640 GTTCTCGCAGGATGTCACTAAGTCTCGCGGACCGTCAATATATATACGTTAGCGGG 15699  
 QY 427 AlaArgValAsnLeuProProValAlaValValSerProGlnLeuGlnGluLeuThr 446  
 DB 15700 CGCAACAGTCCGGA---ACGCAACACGAGCATTACCCATCCG-----CTGACT 15744  
 QY 447 LeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspAsp----- 462  
 DB 15745 GTTGATCTCACCAGCGTCCCATCAGCATGAACAGCATCACAGCGATGATCGATTAAAC 15804  
 QY 462 ----- 462  
 DB 15805 GCCGCCGAAAAAGCGCGCTTAACGCTCTCCGCGCAGCAGCTCCGGCGTCCGAGCGGG 15864  
 QY 463 ---ThrGluIleValSerTyrHisTrpGlu-----GluIleAsnGly 475  
 DB 15865 CAAACCGCTCACCCTTGGCGGCAAACTACACACACCGTGGCGGCAACCGT 15924  
 QY 476 ProPheIleGluGluLysThrSerValAspSerProValLeuArgLeuSerAsnLeuAsp 495  
 DB 15925 AGCTGG-----AGCACACCGTTCGGCGGGGATTTGGCGGCTTGGCT 15969  
 QY 496 ProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThr 515  
 DB 15970 GATGCGGATGCCAGCGCCAGGTGGGTGACTTAACGTCAACGGC-----AACAGCGCC 16023  
 QY 516 ThrAlaAla-----LeuIleValAsnAlaVal 525  
 DB 16024 ACAGCAACGACGAATACAGCGTCGATAGCCCGCGCCACCGTGACCATCAACACCAT 16083  
 QY 526 AspTyrProProValAlaAsnAla-----GlyPro 535  
 DB 16084 GCCAGCGTAAATCATCATCAGCCAGCAGCGGCGGGGGGTAAACGGTGTCCGCGCAC 16143  
 QY 536 AsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAsp 555  
 DB 16144 AGCACCGCGCAACCGCGCAGACGCTCACCGTCACTTAACGGCACT----- 16191  
 QY 556 AspHisGlnIleValLeuTyr-----GluTrpSerLeuGlyProGlySerGluGly 572  
 DB 16192 AACTACCAGACCGGTGCAGACAGCAGCGAGCTGGAGTTTA----- 16233  
 QY 573 LysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGlu 592  
 DB 16234 -----ACGTGCGCGCCAGCGACTTACCGCACTTCTGCTAAT 16269  
 QY 593 GlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla\*\*\* 612  
 DB 16270 AACGCTACACCTGACCGCCAGCGTCAGCGATCTGGGGGTAACTTTGGCGAGCGCCAGC 16329  
 QY 613 ValThrValIleValGlnProGluAsnAsnArgProProValAla-----ValAla 629  
 DB 16330 AAAGCGGTGACCGTC-----GATACCACTGCGCGGTGATGAGTTTAAACACCGTGGCG 16383  
 QY 630 GlyProAsp-----LysGluLeuIlePheProValGluSerAlaThrLeu 644

```
Db 16384 GCGCATGATGTGATTAAACAACGTGGAACACATTTCAGGCGCAGATATTATTCAGCGCCACCGCC 16443
Qy 645 AspGlySerSerSerSerAspHisGlyIleValPheTyrHisTrpGluHisValArg 664
Db 16444 ACGGGCGGTGGCGGCGCAGCCCGCTGGTGGTGCATCCGCGGCGAGCATGTATGTACC 16503
Qy 665 GlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeu 684
Db 16504 AGCACCGATGCCAGCGGCAACTGGAGCGTGGCGTCCCTGCCAGCGTGATTTCCGGCTT 16563
Qy 685 GlnValGlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThr 704
Db 16564 GCGGATGGACCGGTGACCATCAGCGCCACCATTTACCGATAGCGCAGCAACAGCAGCAG 16623
Qy 705 SerThr-----LeuThrValAlaValLysLys 713
Db 16624 CAGACCAACAAGTACAGGTGAACGCGGCGAGTGCCTCTCGTGCAGCACTATTCAGC 16683
Qy 714 GluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuValLeuProAsn 733
Db 16684 GCGGATAACCTTATTAAACGCCCGGAAGCGGSC----- 16716
Qy 734 AsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeu 753
Db 16717 AGTGGCTGACCTGACCGGCACT----- 16740
Qy 754 TrpIleArgAspGlnSerProAlaAlaGlyAspValIle-----AspGly 769
Db 16741 -----GGCACTAATTTCGCGACAGGTACAGTGTGTCGCTGTGCTTACCGC 16788
Qy 770 SerAspHisSerValAlaLeuGln----- 777
Db 16789 AAAGGCTACAGCGCCACCATTCAGACCAACGGGAGCTGGAGCGTGAACGTGCTCGCGG 16848
Qy 778 ----LeuThrAsnLeuValGluGlyVal----TyrThrPheHisLeuArgValThrAspSer 795
Db 16849 GATGTTGGCGCACTAGTGTATGCGCAGCTACCGTACCGGTTAGCGCTCCGCTCAGGACAGT 16908
Qy 796 GlnGlyAlaSerAspThrAspThrAlaThrValGluVal 808
Db 16909 GCGGAAACAGCGCCACCGCTCGCGCAGCGTGGCGGTG 16947

RESULT 11
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)

; LOCATION: (84808)..(84808)
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600592)..(600592)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
```

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g

```

```

US-08-916-421B-1
Alignment Scores:
Pred. NO.: 0.00462 Length: 1664976
Score: 201.00 Matches: 128
Percent Similarity: 35.63% Conservative: 74
Best Local Similarity: 22.57% Mismatches: 217
Query Match: 3.60% Indels: 149
DB: 4 Gaps: 26

US-10-764-390-3 (1-1072) x US-08-916-421B-1 (1-1664976)
Qy 298 GlySerThrGluHisSerIleProThrProThrSerAlaAlaProSerGluSerThr 317
Db 33829 GGAATATAGAGTTGGAAATTC---CCCAATTGGAGTTGGAGGATACAGCGTTAT 33773
Qy 318 ProSerGluLeuProIleSerProThrAlaProArgThrValLysGluLeuThrVal 337
Db 33772 CCT---CAAAATTTATGCTTCTCTACAATAGCCCT---ACTTACTCCAATAAATATA 33719
Qy 338 SerAlaGlyAspAsnLeuIleThrLeuProAspAsnGluValGluLeuLysAlaPhe 357
Db 33718 ACA-----ATA 33713
Qy 358 ValAlaProAlaProProValGluThrThrTyrAsnTyrGluTyrAsnLeuIleSerHis 377
Db 33712 GTAGAGCGGCA-----TCGTGACTTGGTATTCGCATCATATA----- 33674
Qy 378 ProThrAspTyrGlnGlyGlu-----IleLys---GlnGlyHisLysGln 391
Db 33673 -----TATTCGGAGAGTGGCGAGTGTTCGGATAAACCCCAAAATGCCCATAT 33623
Qy 392 ThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSer 411
Db 33622 ACTTTCACATACATACACTTCAGAAAGGGTTTATCCAAATATATGAAAGTTTATACT 33563
Qy 412 GluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgValAsn 431
Db 33562 GCAGAAATATGAAACAGTATATGTAATAGACAAATAAAACCCCAACTAGCAACTCTAT 33503
Qy 432 LeuProProValAlaValValSerPro-GlnLeuGlnGluLeuThrLeuProLeu----- 449
Db 33502 ATTTATCCA-----AATCTCGCAAGTTATAAGATACAAATTCCTTTAGTCCA 33455
Qy 450 -----ThrSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGlu-- 464
Db 33454 ATAAATAGTTATGACCCAGATGCTAATAGATGATTCAGAGTATAATTACTACGGCACC 33395
Qy 465 -----IleValSerTyrHisTrpGluGlu----- 472
Db 33394 TTAATAGGGTATTATCCTATATCTCCAAACTCACCCATGGCAGAATATATGGATTAAAC 33335
Qy 473 -----IleAsnGlyProPheIleGluGluLysThrSe 483
Db 33334 TTAACAGTTTATGATTCTTAATGGAAATGTCATGGGAAT-TACTCATCAATGAATTAAC 33276
Qy 483 rValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLe 503
Db 33275 TATTATATCC-----AAAAGTTTCTTATAGGGAATTACACTGCTAAATTT 33231
Qy 503 uThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuIleValAsnAs 523
Db 33230 GGTTAGTTTGGGATGGAATGGGAGGAGTTAATTCAACAACCTGTTAAGTTTAGTGTG----- 33176
Qy 523 nAlaValAspTyrProProValAla-----AsnAlaGlyProAs 536
Db 33175 -----ATAAATAGACCTCTGTGGCTCAATTTATATCTATCCAGATAAACCTGAACTTAA 33120
Qy 536 nHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAs 556
Db 33119 TGAGGATCTAGAAATTCGTTTCCAGTCTTATGACCCGGAGGGT----- 33077
Qy 556 pHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyHisValVa 576

```

## RESULT 12



```

; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312933)..(312933)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature

```

	:	LOCATION:	(779676)..(779676)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(855539)..(855539)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(871619)..(871619)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(1084830)..(1084830)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(1096846)..(1096846)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(1119881)..(1119881)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(1130881)..(1130881)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(1310988)..(1310988)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(1313224)..(1313224)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	FEATURE:		
	:	NAME/KEY:	misc feature	
	:	LOCATION:	(1349473)..(1349473)	
	:	OTHER INFORMATION:	n equals a, t, c, or g	
	:	Alignment Scores:		
	:	Pred. No.:	0.00462	Length
	:	Score:	201.00	Matches
	:	Percent Similarity:	35.63%	Conserv.
	:	Best Local Similarity:	22.57%	Mismat.
	:	Query Match:	3.60%	Indels
	:	DB:	4	Gaps
	:	US-10-764-390-3 (1-1072) x US-09-692-570-1 (1-		
	:	Qy	298 GlySerThrGluHisSerIleProThrProp	
	:	Dd	33829 GAAATATAGAGTTGGGAATTCCA---CCAC	
	:	Qy	318 ProSexGluLeuProileSerProThrThra	
	:	Dd	33772 CCT--CAAATTTATCTTCTCCTACAATAG	
	:	Qy	338 SerAlaGlyAspAsnLeullelthrLeuP	
	:	Dd	33718 ACA-----	
	:	Qy	358 ValAlaproAlaproValGluThrThry	
	:	Dd	33712 GTAGACCGCA-----TCGT	
	:	Qy	378 ProThrAspyrrGlnglyGu---	
	:	Dd	33673 -----TAITTCGAGATGTTGCAGTG	
	:	Qy	392 ThrLeuAsnLeuSerGlnLeuSerValGlyL	
	:	Dd	33622 ACTTTCACATCATACACTTCAGAAGGGG	

```

1664976
128
Live: 74
ss: 217
26 149
26
664976)
ThrSerAlaalaProSerGluSerThr 317
ATTGGAGTTGGAGGATACAGCGTTAT 33773
ProArgThrValLysGluLeuThrVal 337
CTTACTATCCAATAAATATA 33719
AspAsnGluValGluLeuLysAlaPhe 357
-----ATA 33713
AsnTyrGluTrpAsnLeuIleSerHis 377
CTTGGTATTGGCATCATATA----- 33674
-----IleLys---GlnGlyHisLysGln 391
TGGATAAACCCAAAAATCGCCATAT 33623
TyrValPheLysValThrValSerSer 411
TATCCAATATATATGAAGTTTATACT 33563

```

QY 412 GluAsnAlaPheGlyGluGlyPheValAsnValThrValValProAlaAlaArgValAsn 431  
 Db 33562 GCAGAAATATGAACACAGTATATGTAATAGACAATAAACCACCAATAGCCTAT 33503  
 QY 432 LeuProProValAlaValValSerPro-GlnLeuGlnGluLeuThrLeuProLeu----- 449  
 Db 33502 ATTATATCCA-----AATCCTGCAAGTTATAAAGATACAAATTTCTTTAGTCCA 33455  
 QY 450 -----ThrSerAlaLeuLeuAspGlySerGlnSerThrAspAspThrGlu-- 464  
 Db 33454 ATAAATAGTTATGACCCAGATGCTAATAGAAAGTATCCAGAGATATAATTACTACGGCACC 33395  
 QY 465 -----IleValSerTyrHisTyrGluGlu----- 472  
 Db 33394 TTAATAGGTATATCTCTATATCTCCAAACTCACCCTGCAAGATATATGATGATTTAAC 33335  
 QY 473 -----IleAsnGlyProPheIleGluGluLysThrSe 483  
 Db 33334 TTAACAGTTTATGATTTCTAATGGAATGTTGCATGGAAT-TACTCATCAATGAATTAAC 33276  
 QY 483 rValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArgIle 503  
 Db 33275 TATTATATCC-----AAAAAGTTTCTCTATAGGAATTAACACTGCTAAAT 33231  
 QY 503 uThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuIleValAsnAs 523  
 Db 33230 GGTAGTTTGGATGGAATGGAGGAGTTAATTCACCAACTGTTAAGTTTGTG----- 33176  
 QY 523 nAlaValAspTyrProProValAla-----AsnAlaGlyProAs 536  
 Db 33175 ----ATAAATAGACCTCTCTGTGGCTCAATTTATATATCTATCAGATATAAACCCTGAACCTAA 33120  
 QY 536 nHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAs 556  
 Db 33119 TGAGGATGAGAAATCGTTTCCAGCTCTTANGACCGAGGGT----- 33077  
 QY 556 pHisGlnIleValLeuTyrGluTyrSerLeuGlyProGlySerGluGlyLysHisValVa 576  
 Db 33076 ----GAAATGTCATATTACATTTGGAATTTGGAGATGGACC-----GT 33036  
 QY 576 lMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrTh 596  
 Db 33035 TATCAATACAAACAGATACG---ATAGTTTCATCAAAATATGAAGACAGGCTCTATATAC 32979  
 QY 596 rPheGlnLeuLysValThrAspSerArgGlnGlnSerThrAla\*\*ValThrValI 616  
 Db 32978 AGTTACTTTAACAGTATTTGACAAATATAACGCAAGTAGTTCAATATCAAAACAATAGT 32919  
 QY 616 eValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLysGluLeuI 636  
 Db 32918 AGTTGGGGT-----ATATTAGCTAACTTTACATATGAATATAT 32880  
 QY 636 e-----PheProValGluSerAlaThrLeuAspGlySerSerSe 650  
 Db 32879 AAATGAAATACGTGAAGATTTGAAGATACATCAGTTGTGGCCCTGGGAAA----- 32828  
 QY 650 rAspAspHisGlyLysValPheTyrHisTyrGluHisValArgGlyProSerAlaValGl 670  
 Db 32827 -----ATAATTAGTTGGCACTGGGACTTTGGAGATGGCAGCATCAACAGA 32781  
 QY 670 uMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHi 690  
 Db 32780 ACAAATCTATTCAT-----ACTTACAGCAAGAGGGGCATCTT 32739  
 QY 690 sPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAl 710  
 Db 32738 TGTAACCTTAACCTGTTAAAGCGCATACAAACTTAACGATAGTGTAAACAAGATTCTTAT 32679  
 QY 710 aValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuVa 730  
 Db 32678 AATTCTTCCACCTCCAAAATATCTCTCAGTTGCTGACTTTACGTATATAA-----GT 32628  
 QY 730 lLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAsp---AspGlnArgI 749

Db 32627 AATAAAATGGAACACTACTGTAGAGTTTAAACGCTAAATTTATCATACGATGAAGATGGATATAT 32568  
 QY 749 eValSerTyrLeuTyrPheIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGl 769  
 Db 32567 TAAATATTATATATCG-----GACTT 32547  
 QY 769 ySerAspHisSerValAlaLeuGlnLeuThrAsnLeuVal----- 782  
 Db 32546 TGGAGATGGAACACACGCAAGATATACAAACAATCCAAATTTATATACACAATATAAAAA 32487  
 QY 783 -GluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAs 802  
 Db 32486 AGAGGGGTTTATCCAGTTTACCTTCACAGTTATAGATAATGATAATCTTTTCAGATTCAAG 32427  
 QY 802 pThrAlaThrValGluVal 808  
 Db 32426 AACTAAGTTAATAAATATA 32408  
 RESULT 13  
 US-09-621-976-13082  
 ; Sequence 13082, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621.976  
 ; CURRENT FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 13082  
 ; LENGTH: 320  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-621-976-13082  
 Alignment Scores:  
 Pred. No.: 1.56e-09 Length: 320  
 Score: 199.00 Matches: 40  
 Percent Similarity: 86.96% Conservative: 0  
 Best Local Similarity: 86.96% Mismatches: 6  
 Query Match: 3.57% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-764-390-3 (1-1072) x US-09-621-976-13082 (1-320)  
 QY 503 LeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuIleValAsn 522  
 Db 180 TTGACTGTAGTAGACTCTGTAGAGTACCACTCTACTCTGCAAACTGCAGTGAAC 239  
 QY 523 AsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuPro 542  
 Db 240 AAAGCTGTGGATTACCCCTGTGGCCAAACGAGGCCCAACCAAGTGATCATCCCTGCC 299  
 QY 543 GlnAsnSerIleThrLeu 548  
 Db 300 CAAAACCTCATCACCCCTC 317  
 RESULT 14  
 US-09-902-540-1136/c  
 ; Sequence 1136, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540



```
Db 5701 CGCGCGCATCGCGACGAGAAACCGCTGTCATGTGGAC----- 5662
Qy 529 oValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAs 549
Db 5661 -----ACACCGCGCGGAGCTACCTCTCTCACCAGCGGCGCGCGCTG-- 5611
Qy 549 nGlyAsnGlnSerSerAspHisGlnIleValLeuThrGluThrProGlu 569
Db 5610 ----CGTTCCTACAGCGTGGATGTGGGATGTGTGTGCTGCTGCTG-- 5566
Qy 569 ySerGluGlyLeuHisValMetGlnGlyValGlnThrProThrLeuHisLeuSerAl 589
Db 5565 -----CCGCTGCGCGTGGAGACCACTGGGTGTCAGGTGACGGA 5528
Qy 589 aMetGlnGlu-----GlyAspThrThrPheGln----- 598
Db 5527 CGTGCCCGCGCGCGCGGAGCGTGACGACCGCGGTGGACCTGTGGCGGGGAGAACCT 5468
Qy 599 -----LeuIysValThrAspSerSerArgGlnGlnSerThrAla**ValThrValIl 616
Db 5467 CATCTGTGTCCGAGCCAGCGATTCGGTAGGCCACGTCACGAGCAGATGCTGACGGTGTG 5408
Qy 616 eValGlnProGluAsnAsnArgProValAlaValAlaGlyProAspLysGluLeuIl 636
Db 5407 GTTGGACGCGAG-----GCACCGTGGGTGTCCACGGGCGCTGCCGGATGGGTGCTCGT 5354
Qy 636 ePheProValGluSerAlaThrLeuAspGlySerSerSerSerAspHisGlyIleVa 656
Db 5353 CGGGCGG-----CAGCGGGGAGAGCGATGCCCTAGCAGATCGGGGTGTA 5309
Qy 656 lPheThrHisTrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAspIy 676
Db 5308 C-----TC 5306
Qy 676 sAlaIleAlaThrValThrGlyLeuGlnValGly---ThrTyr----- 689
Db 5305 CGCGTCCGCCACACACGCGTACGCTGTCTCACGGGGCGGACGTACACGCTCCCGGGGTGG 5246
Qy 690 -----HisPheArgLeuThrVa 695
Db 5245 CGCGCGCGTGCAGCGACGCTGACCTGTGTGCGCGGGGACACGCGTTACCATCGACGT 5186
Qy 695 lLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysGluAs 715
Db 5185 GACGAGCGAGACGGCATGTGCGGCCACGCTGTGCGCGACGGTG-----CGCTACGA 5135
Qy 715 nAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAsnSe 735
Db 5134 CGACGCGCGCGCGCGCTCA-----CTGCTGCTCGCGTGCCTGCGCGG 5093
Qy 735 rIleThrLeuAspGly-----SerArgSerThrAspAspGlnArgIleVa 750
Db 5092 GCAGAGTACAGCGGCGTGTATTCATCATCACCAGCGGGGTGACGGAC-----CGCGTGAG 5039
Qy 750 lSerThrLeuThrIle-----ArgAspGlyGlnSerProAlaAlaGlyAspValIl 767
Db 5038 CGCGCGTGGTGGTGGTGGCTTCACCGCGGATGCTCCGGCATCCGGCGCGCCACGCTCCA 4979
Qy 767 eAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValThrTh 787
Db 4978 GCCGAGCGGCACCTGTGTGGCGGAGCTGGACACGCGCGAGCTGGTGGCGGGGAGCACAC 4919
Qy 787 rPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAspThr----- 803
Db 4918 CTGTGGAGTGTGTGATGATGATGACCGCGGGGGAACCTTCGCCATCCAGACGTTCCCTTCGT 4859
Qy 804 -----AlaThrValGluValGlnProAspProArgLysSerGly 816
Db 4858 CACGCGGAACACCGCTAGGCTACTTCTTCTTGTGTGTGATCGCCGAGCGGATCAGGT 4801
RESULT 15
US-08-853-659A-5
; Sequence 5, Application US/08853659A
```

```
; Patent No. 5925522
; GENERAL INFORMATION:
; APPLICANT: Wong, K.K.; Saffer, J.D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods Of Detection
; TITLE OF INVENTION: Of A
; TITLE OF INVENTION: Salmonella Sequence, And Methods Of Detection Of Salmonella
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Paul W. Zimmerman
; ADDRESSEE: Intellectual Property Services
; ADDRESSEE: Battelle Memorial Institute
; ADDRESSEE: PNNL P.O. Box 999
; STREET: Washington Way
; CITY: Richland
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 99352
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (WordPerfect 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/853.659A
; FILING DATE: Unknown
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE: n/a
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15512 bases
; TYPE: nucleotide
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; US-08-853-659A-5

Alignment Scores:
Pred. No.: 3.13e-06 Length: 15512
Score: 196.00 Matches: 164
Percent Similarity: 23.17% Conservative: 100
Best Local Similarity: 20.60% Mismatches: 243
Query Match: 3.51% Indels: 290
DB: 2 Gaps: 38

US-10-764-390-3 (1-1072) x US-08-853-659A-5 (1-15512)

Qy 271 ProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGlu----- 288
Db 13278 CCGACGTTTTTACTGGGCAATATTGACGACGACGCGCGGTATGTCCAGGTTGAGGTGCAA 13337
Qy 289 -----LysSerProValLeuThrValThrProGlySerThrGluHisSerIlePro 305
Db 13338 CATGCGCGGACGAAAGAAGTGTCTGCGGCCACCAAGGCGGACCGGCATCTGGAGCGTG 13397
Qy 306 ThrProThrSerAlaAlaProSerGluSerThr-----ProSerGlu 320
Db 13398 ACACCGACCGGCACATGGCAGATGGCGACTATACGCTGACCGTGCAGGTGAGGATGAT 13457
Qy 321 LeuProIleSerProThrThrAlaProArgThrVal-----LysGluLeuThrVal 337
Db 13458 GCGGGGAACGTAAATACTACGCGCGCTGACGGTTCAGCGTTTGACACCAATCATCATC 13517
Qy 338 SerAla-----GlyAspAsnLeu----- 344
Db 13518 GATGTTATTGAAGTGTAAATGATAACGGTATTCCTCCGGCGGACCACTGACCAATGACGTT 13577
Qy 345 -----IleThrLeuProAsp-----AsnGluValGluLeu----- 354
Db 13578 CGTCCACACTTCCGGGTACGCGTGCAGGGGATGTCAACGAAGTACGTCTGAGTATCGAC 13637
Qy 355 -----LysAlaPheValAlaProAlaProValGluThrThrTyrAsnTyrGluTrp 372
```

Db 13638 GCGGTATACGTGGTTCGTGCAACACAGGCGCGGAGGATCTGGATTACACCTGG 13697  
Qy 373 -----AsnLeuIleSerHisProThrAspTyrGlnGly 383  
Db 13698 CCGAAGATGTGACCGCGGTACATACCTGACGGTAGAAGCGACCGAT----- 13748  
Qy 384 GluIleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyr 403  
Db 13749 -----AAGCGGGAATATAACGACGCGAGCGCTCGAT----- 13781  
Qy 404 ValPheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThr 423  
Db 13782 -----TTTACCATTTGATACC----- 13796  
Qy 424 ValLysProAlaArgValAsnLeuProProValAlaValValSerProGlnLeuGln 443  
Db 13797 -----CGGCTGTCAACGCTACCATCGCTATG----- 13823  
Qy 444 GluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspAspThr 463  
Db 13824 -----GATAGCAGGAGCATACA 13841  
Qy 464 GluIleValSerTyrHisTrpGluIleAsnGlyProPheIleGluGluLysThrSer 483  
Db 13842 GGTGCCATTGGCATCAT-----ATTACGAGC 13868  
Qy 484 ValAspSerProValLeuArgLeuSerAsnLeuAsp----- 495  
Db 13869 GTCAAAAGACCGGGCTTTACTATTGGCAATATTGACGCGGATGCGCACTCGTCAATTG 13928  
Qy 495 ----- 495  
Db 13929 CCGATCACACAGGCGGCAANTAGCCAGGAAGTGACACTAACCCAGTTGGAGGACAGTGG 13988  
Qy 496 -----ProGlyAsnTyrSerPheArgLeuThrValThr 506  
Db 13989 CGCTTTACCGCAGATGCTGACTGGCGGACGCTAGCTATACGCTGACGGTAGAGTTAAG 14048  
Qy 507 AspSer-----AspGlyAlaThr 512  
Db 14049 GATAACGCGAGAAACGTTTCGTGAGTCCACGCGCTGGTGGTGACGGTGACACGCAAC 14108  
Qy 513 AsnSerThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProProValAlaAsn 532  
Db 14109 AGCATTAAGTATTAATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 14168  
Qy 533 Ala--GlyProAsnHisThrIleThrLeuProGln-----AsnSerIleThrLeuAsn 549  
Db 14169 AGCACCCGTCGCGAGTTGAGATCAACGCTGCGCGGAGTGAATTCGTGCAACTGAGC 14228  
Qy 550 GlyAsnGlnSerSerAspHisGlnIleValLeuTyrGluTrp---SerLeuGlyPro 568  
Db 14229 ATTGATGGGGCGCAAC-----TGGGTGAGCGCGCGCAG 14264  
Qy 569 GlySerGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSer 588  
Db 14265 GGTATCGAAGG-----GTCTGGGGCTATACCTGGCCA-----ACG 14300  
Qy 589 AlaMetGlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGln 608  
Db 14301 GATATGGCGGATGGAACACACACCTTAACCGCTGATGATGATGATGATGATGATGATGAT 14360  
Qy 609 SerThrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaVal 628  
Db 14361 GCGAGCAACCGTTGAATTTTCATCGACACCGGTTGTCGACCGCCACCAATGCGGTG 14420  
Qy 629 AlaGlyProAspLysGluLeuPheProValGluSerAlaThrLeuAspGlySerSer 648  
Db 14421 -----GAT 14423  
Qy 649 SerSerAspAspHisGlyIleValPheTyrHisTrpGluHisValArgGlyProSerAla 668  
Db 14424 AGCAGGATGATACCGGTACGCTGGCGATGATGATGATGATGATGATGATGATGATGATGAT 14483

Qy 669 ValGluMetGluAsnIleAspLysAlaIleAlaThrValThr----- 682  
Db 14484 ATT---CTGCAGAAATATCGATTTCGATGTTTATCAACGTTACAGTCAGCGTACGCATAAT 14540  
Qy 683 GlyLeuGlnVal----- 686  
Db 14541 GGAACGACAACTCTGTTACTCGACACAGGGGCTGGAGATGGAGTTCACCGGCCA 14600  
Qy 687 -----GlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGlyLeu 701  
Db 14601 GCGCGGTGGCGGACGCTGATTATACGCTGACGCTGACAGTGGAGGATCGGCGGGAAT 14660  
Qy 702 SerSerThrSerThr---LeuThrValAlaValLysLysGlu--- 714  
Db 14661 ACGCTCGCTCTACGCGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 14720  
Qy 715 -----AsnAsnSerProArgAlaArgAlaGlyGlyArgHisVal----- 728  
Db 14721 GAATTAGTCAACGATACGCGGCTCCCTGGCGATTAATGTGACAAACATGTGCTCCGCGAG 14780  
Qy 729 -----LeuValLeuProAsnAsnSer-----IleThrLeuAspGlySerArg 742  
Db 14781 TTCCAGATCTCGGTACCGGATGATGT--GGAAAAGTCTCTCTGAGTATTGACGCGGACG 14839  
Qy 743 SerThrAspAspGlnArgIleValSerTyrLeuTrpIleArgAspGlyGlnSerProAla 762  
Db 14840 ACC-----TGGGTACTGCAATCAAGATTCGACG 14869  
Qy 763 AlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuVal 782  
Db 14870 GCTGCG---ATTGGGATTACAC-----TGGCCGACGGATATGCCA 14908  
Qy 783 GluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAsp 802  
Db 14909 GAGGACAGCATACCTGACCGGTGAAGTGAATGACTGACGCTGGGTGTAATAAGATGACGAG 14968  
Qy 803 ThrAla-----ThrValGluValGlnProAspProArgLysSerGlyLeuValGluLeu 820  
Db 14969 ACGCTCAATTTCATCTATCGATATCACG-----TTGTTAACGCCA 15007  
Qy 821 ThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAsp-----ThrLeu 837  
Db 15008 ACCATTGAGCTAGCGCGCGATCAGGATACCGGACAGATAGAACGATAATCTGACCGAGC 15067  
Qy 838 ValArgGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLysValGlnLysIle 857  
Db 15068 GTCACTCAGCGGTATTGTTGTTGGGAGTATCGATAAAGATGTTTGA----- 15115  
Qy 858 ArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSerArgProPhe 877  
Db 15116 -----CACGTGGAATTGAGT-----ATTGAGCATACGCGACGTTT 15151  
Qy 878 LysValLeuLysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerLysGluLys 897  
Db 15152 AAAACGGGTGACTACCGCAATCAGCGACGCGTGGCGCTATCGACCGGATTCGCTTTG 15211  
Qy 898 AlaAsp---PheLeuLeuPheLysValLeuArgValAspThrAlaGly 912  
Db 15212 GCGAGCGGTAGCTACATTCACCGTACGCTAACAGATGTGGCAGGC 15259

Search completed: October 12, 2005, 16:36:49  
Job time : 3747 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 12, 2005, 10:31:17 ; Search time 6860 Seconds  
(without alignments)  
5948.236 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSLLLVITAGC.....VSMNGSRNGASFSYCSKDR 1072

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2 1/USPTO.spool p/US10764390/runat 12102005 110203 22340/app query.fasta\_1.1223  
-DB=EST -OFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCF=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10764390.@CGN 1 1 5180 @runat 12102005 110203 22340 -NCPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hc.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gsl.\*  
9: gb\_gsl2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4006.5	71.8	4828	3 AK051381	AK051381 Mus muscu
2	2420.5	43.4	3660	3 AK084668	AK084668 Mus muscu
3	2420.5	43.4	4141	3 AK043006	AK043006 Mus muscu
4	2389.5	42.8	3616	3 AF289597	AF289597 Homo sapi
5	2310	41.4	3067	3 AK031342	AK031342 Mus muscu
6	2256.5	40.4	3671	3 HSM805357	AL834315 Homo sapi
7	1982	35.5	2959	9 AY417695	AY417695 Mus muscu
8	1940.5	34.8	2962	9 AY417693	AY417693 Homo sapi
9	1635	29.3	2962	9 AY417694	AY417694 Pan trogl

10	1471	26.4	2189	3	AK049570	Mus muscu
11	1312	23.5	842	1	AU130115	AU130115
12	1051	18.8	710	6	CB722905	UI-M-GH0-
13	1025	18.4	597	1	AL044313	DKFZp4340
14	999	17.9	598	5	BP229849	BP229849
15	994	17.8	842	7	CK463785	CK463785 934714 MA
16	922	16.5	750	7	CN527744	UI-M-H00-
17	916	16.4	534	7	CR536629	CR536629 DKFZp459A
18	910	16.3	927	5	BQ897270	BQ897270 AGENCOURT
19	890	15.9	821	1	AU129983	AU129983
20	877	15.7	814	1	AU131313	AU131313
21	870	15.6	815	7	CR566736	CR566736
22	862	15.4	610	6	CB519102	CB519102 UI-M-GH0-
23	849	15.2	875	6	CA790104	CA790104 AGENCOURT
24	845	15.1	772	1	AU132238	AU132238
25	841.5	15.1	909	5	BQ422326	BQ422326 603019005
26	836	15.0	494	5	BQ267742	BQ267742 i93e06.Y
27	820	14.7	784	4	BF166063	BF166063 603053031
28	818	14.7	1056	4	BM547013	BM547013 AGENCOURT
29	816	14.6	685	6	CB527694	CB527694 UI-M-FY0-
30	812	14.6	727	5	BQ768962	BQ768962 UI-M-FC0-
31	800	14.3	1107	4	BM479855	BM479855 AGENCOURT
32	792	14.2	479	1	AL589494	AL589494 DKFZp451G
33	789.5	14.1	1040	2	BE311921	BE311921 601143520
34	784.5	14.1	949	4	BF981495	BF981495 602309331
35	784	14.1	653	4	BI390119	BI390119 pgpic.BK0
36	763.5	13.7	805	4	BI159118	BI159118 602922880
37	761	13.6	663	7	CN365571	CN365571 170004245
38	757	13.6	655	7	CF899230	CF899230 A0300F03-
39	753	13.5	777	7	CR765174	CR765174 DKFZp468E
40	752	13.5	638	1	AL858156	AL858156
41	750	13.4	964	4	BG172757	BG172757 602337184
42	749.5	13.4	968	2	BE617479	BE617479 601442136
43	738	13.2	829	7	CF579859	CF579859 AGENCOURT
44	737	13.2	616	7	CN704544	CN704544 E0489B01-
45	737	13.2	617	7	CF170107	CF170107 B0822F05-

#### ALIGNMENTS

AK051381 4828 bp mRNA linear HTC 03-APR-2004  
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length  
enriched library, clone:DI30043K22 product:hypothetical PKD domain  
containing protein, full insert sequence.

AK051381 1 GI:26094501

VERSION  
HTC; CAP trapper.

KEYWORDS  
HTC; CAP trapper.

SOURCE  
Mus musculus

ORGANISM  
Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;

REFERENCE  
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

AUTHORS  
Carninci, P. and Hayashizaki, Y.

TITLE  
High-efficiency full-length cDNA cloning

JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)

MEDLINE  
99279253

PUBMED  
10349636

REFERENCE  
2

AUTHORS  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL  
MEDLINE  
20499374

PUBMED  
11042159

REFERENCE  
3

AUTHORS  
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 4828)

JOURNAL  
REFERENCE  
AUTHORS

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishi,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submission

JOURNAL  
REFERENCE  
AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.jp,  
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.jp/  
URL:http://fantom.gsc.riken.jp/

FEATURES

source

1. .4828  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:D130043K22"  
/db\_xref="taxon:10090"  
/clone="D130043K22"  
/issue\_type="spinal ganglion"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
258..3364  
misc\_feature  
/notes="hypothetical PKD domain containing protein  
(InterPro|IPR000601, evidence: InterPro)  
putative"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	4828
Score:	4006.50	Matches:	779
Percent Similarity:	80.41%	Conservative:	91
Best Local Similarity:	72.00%	Mismatches:	155
Query Match:	71.80%	Indels:	5
DB:	3	Gaps:	

US-10-764-390-3 (1-1072) x AK051381 (1-4828)

QY	1	MetAlaProThrGlyValLeuSerSerLeuLeuLeuValThrIleAlaGlyCys	20
DB	258	ATGTTGTCCTCCACAGAGTACTCTCTTCATTGCTGCTGGCAGCATGGCAGAGGC	317
QY	21	AlaArgLysGlnCysSerGluGlyArgThrPyrSerAsnAlaValIleSerProAsnLeu	40
DB	318	AGTTCTCAGCAGTGTCTCTGAGGCGCAGGACTTACTCCGATGTCATCTTACCTAACCCG	377
QY	41	GluThrThrArgIleMetArgValSerHisThrPheProValValAspCysThrAla	60
DB	378	GAACCATCATGATGCGGGTGTCTCAAACTTCTCCGTTGGAGACTGCACGCGCGCT	437
QY	61	CysCysAspLeuSerSerCysAspLeuAlaTrpPheGluGlyArgCysTyrLeuVal	80
DB	438	TGCTGTGACTTGCTACCTGTGACTGGCTGTGGTGTGGGGCAGCTGCTATCTGTG	497
QY	81	SerCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTyrLeu	100
DB	498	AAATGTCATGCGCTCGGAGAAATTGGAGCCAGGACACAGGCCCATCCGGTCTTACCTC	557
QY	101	ThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMet	120
DB	558	ACTTTCGTGCGCAGACCTGTCCAGAGGCCCGGGCAGCTGCTGGACTATGGAGACATGATG	617
QY	121	LeuAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArgLysAsp	140
DB	618	CTGAGCAGGCGCTCCCTCTCAGAGCTTGGGGAGACTCCCTTGGAGACTCAGGAGGAC	677
QY	141	Leu***PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTyr***AspAspTyr	160
DB	678	TTGCCCTTTCTTGGCAAGACGGGGGACACAGAGGAGACCATGAGTACTCAGATCAATAC	737
QY	151	ArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAla	180
DB	738	AAAGACCTGGAGCGGGCGCTCTCGAGCCCTCCAAACAGAGAGCCCTTAGAGGTAGCGCA	797
QY	181	GluTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsn---SerSerVal	199
DB	798	GAGTACCCGCTGGAGCTCTGCTCCAGCAGCAACGAGGAGGTTTCAATGCTACAGCTACA	857
QY	200	GlyAspSerProAlaValProAlaGluThrGlnGlnAspPro-----	213
DB	858	GGAGACAACCTCAGCTGCTTCCATGGAGAAGCTGCAGGATCCACGCCCCACCCACCTGGAC	917
QY	214	-----GluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeuProGlu	231
DB	918	CAGGAGCAGCTGCAGGCGCTGAATGAGTCGACTTGGTCCCTTACACCCCGGCGACTTCTCA	977
QY	232	ArgSerValLeuLeuProLeuProThrThrPro---SerSerGlyGluValLeuLys	250
DB	978	ATAGCAGTGTGGCTTCTCTCGAGAGCCATTAATCTACAGAGAGGAGCTGGAAGA	1037
QY	251	GluLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLysGluValLeuMet	270
DB	1038	GAAGAGACTTTTACAGCTCCAAAGAACACCGAGCAATAGCTCTGGAAGAGAGGTTCCAATG	1097
QY	271	ProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGluLysSer	290
DB	1098	CTTCTCCATAATCTCTCCCTGCGAGCTGAAGTCTAGCCAGCCAGCCAGCAAGAAAAAAC	1157
QY	291	ProValLeuThrValThrProGlySerThrGluHisSerIleProThrProThrSer	310
DB	1158	TCCAACTTTACAGTCACGCCACGGAGCAGAAAAACACAGACCCCACTTTTCTACCACT	1217
QY	311	AlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaProArg	330
DB	1218	ACAGCTCTCACTGGGCTAAACCCCTCTCGTGGCCCTTGTACCT---ACTGTTCCAGG	1274
QY	331	ThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleThrLeuProAsn	350
DB	1275	ACAGTAAAGGGCGCTTGTGTGTCTGGAGATAACCTAGTACTAATCTTACCGACAGA	1334

Qy 351 GluValGluLeuLysAlaPheValAlaProProValGluThrThrTyrAsnTyr 370  
Db 1335 GAACAGAACTGAAGGCTCTGTGAACACGCGCCCTGCAGATACAACTTACTCTAT 1394  
Qy 371 GluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLys 390  
Db 1395 GAATGGAGTTAATGAGCCACCCAGTAGACTTCCAGGTAAATCAACAGAAACACAG 1454  
Qy 391 GlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSer 410  
Db 1455 CCGACTCTTCACTCTCTCAATTATCTGTGGAGCTCTATGCTTCCAGAGTGGCTTCT 1514  
Qy 411 SerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgVal 430  
Db 1515 AGTGAGAACTGATTTGAGAGGCTGTCAATGTCCAGGTATGCGCAGCTGCAGAGTC 1574  
Qy 431 AsnLeuProProValAlaValValSerProGlnLeuGlnLeuThrLeuProLeuThr 450  
Db 1575 AACCAGCCACTGTAGCTGTGTGTTCTCCCGACACAGAGCTCAGTGTGCTTTGACC 1634  
Qy 451 SerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTrp 470  
Db 1635 TCAGCCCTGATCGATGGCAGCAAGTACAGATGATGATACTGAAATAGTGAGTTACCACTGG 1694  
Qy 471 GluGluIleAsnGlyProPheIleGluLysThrSerValAspSerProValLeuArg 490  
Db 1695 GAAGAAGTCGACGGGCGCTTCTCTAGGAGAGGATTTCCAGCTGACACCCCAATCTTACGA 1754  
Qy 491 LeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGly 510  
Db 1755 TTGTGCAACCTAGTTCCTGTGTTACTACCTTTAGTTGACCATCANGACTCGGATGGC 1814  
Qy 511 AlaThrAsnSerThrThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProProVal 530  
Db 1815 GCTACCAACTCCACCACTGCATCCCTGCTCATCGGTGATGCGGTGACTACCCACCGGTT 1874  
Qy 531 AlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGly 550  
Db 1875 GCCAATGAGGCGCCAAATCAGACCATCACTTTGCCCAAAACACCATCATTTCAATGGG 1934  
Qy 551 AsnGlnSerSerAspAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySer 570  
Db 1935 AACCCAGAGCATGATGATCACAGATGTTCTCTATGATGTTGCGAGGCGCTTGGCGGT 1994  
Qy 571 GluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMet 590  
Db 1995 GAGACAAAGAGATGGTCAATCAGGAGCGCAGACCCCATACCTTCACTTCTGTGAGCTG 2054  
Qy 591 GlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnSerThr 610  
Db 2055 CAGGAAGGAGAGTACACATTTTCAGCTGATGCTGAGCGGATTCCTCGGACAGCAATCCACC 2114  
Qy 611 Ala\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaGly 630  
Db 2115 GCITTTGGTCTGTGACTGTCCAGCTGAGAACCAACAGGCTCCGGTGGCGGTGGCAGGC 2174  
Qy 631 ProAspLysGluLeuIlePheProValGlnSerAlaThrLeuAspGlySerSerSerSer 650  
Db 2175 CCCGACAGAGAGCTGTTTCCCGGTGAGAGTGTCTACACTGGATGGAGCCGAGCAGT 2234  
Qy 651 AspAspHisGlyIleValPheTyrHisTrpGluHisValArgGlyProSerAlaValGlu 670  
Db 2235 GACGATCACCGCATTTGCTGTCTATCACTGCGGAGCACATCAGAGGCCCGCAGTGCAGTGGAG 2294  
Qy 671 MetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHis 690  
Db 2295 ATGGAGATGTGGAAAGCCATAGCCACTGTCCCGGCTTCCAGGTGGGTATCTTACCAC 2354  
Qy 691 PheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAla 710  
Db 2355 TTCCGGTTAACCGTGAGAGACAGCAGGGTCTGAGCAGCAGTCCACCTCAGCTGTGGCT 2414  
Qy 711 ValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyArgHisValLeuVal 730

Db 2415 GTGAAGAGAAATAAACCCCTCCAGAGCCAGGCTGGTGACACATGTTCTTATA 2474  
Qy 731 LeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleVal 750  
Db 2475 CTTCCCAATAATTCCTACTTCTGGATGGTTCAAGGCTCTACTGATGACCCGAGGAATGTG 2534  
Qy 751 SerTyrLeuTrpIleArgAspGlyGlnSerProAlaIleAspValIleAspGlySer 770  
Db 2535 TCCTATCTGTGGATCCCGGATGCCAGAGTCCAGCTCCCGAAC----- 2578  
Qy 771 AspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeu 790  
Db 2578 ----- 2578  
Qy 791 ArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnPro 810  
Db 2578 ----- 2578  
Qy 811 AspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlnLeuThr 830  
Db 2579 ---CCCAAGAGAGATGGCTGGTGGAGTTGATCCTGCAGGTGGTGTGAGCAGCTGACA 2635  
Qy 831 GluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSer 850  
Db 2635 GAGCAGCAGAAAGAGACTCTAGTGAGGACAGCTGGTGTCTCTGAAACGTGCTAGACTCA 2695  
Qy 851 AspIleLysValGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyr 870  
Db 2695 GACGTGAAGGTGTTGAAGATCCAGGCTCACACGGATGTGACACTGTGATTGTTATTAT 2755  
Qy 871 ValGlnSerArgProPheLysValLeuLysAlaAlaGluValAlaArgAsnLeuHis 890  
Db 2756 GTACAGAGTGGATCACCTTTCAAGGTTCTCAGAGCTGCTGCTGGGCCGCGAATCTGCAC 2815  
Qy 891 MetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspThr 910  
Db 2816 AAGCGGCTTTCCAAAGGAAGAGGCTTTCTGCTTTTCAAGGTTCTGAGGTTAGACACA 2875  
Qy 911 AlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCys 930  
Db 2876 GCAGGGTGCCTTCTCAAGTGTTCAGGCCACGGTCACTGTGACCCCATCACAGCGCTGC 2935  
Qy 931 IleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGluSer 950  
Db 2936 ATCTGCTCCAGCTGTGGATGGAGAACCTCATACAGCGGTACATGTGGGATGGAGAGAGC 2995  
Qy 951 AsnCysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuThr 970  
Db 2996 AACTGTGAGTGGAGTGTATTCTATGTGGCAGCTCTGGCATTGACACTAACACTGCTAACT 3055  
Qy 971 GlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLysArgThrLysIleArg 990  
Db 3056 GGAGCTGTGCTGGCTGTGCATCTGCTGTGCAGGAGACGGAAGAGACTAAGTCCGG 3115  
Qy 991 LysLysThrLysTyrThrIleLeuAspAsnMetAspGluGlnGluArgMetGluLeuArg 1010  
Db 3116 AAGAAAACCAAGTATACATCCTAGACAGCATGGACGAAACAGGAAGAAGTGGAACTGAGG 3175  
Qy 1011 ProlysTyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGlu 1030  
Db 3176 CCCAAGTACCGGTATCAAGCAGGAGCAGGAGCAGCAACTCCAGTCTGATGGTGTGGAG 3235  
Qy 1031 SerGluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAsn 1050  
Db 3236 TCTGAGTTTGAACGACACAGGACACGCTATTACGCCGAGAGAGGATGGAGAGGAGTC 3295  
Qy 1051 ProlysValSerMetAsnGlySerIleArgAsnGlyAlaSerPheSerTyrCysSerLys 1070  
Db 3296 CTGAAGGCTTCCCTGAATGGCTGTGCCAGAAATGGAGTTTCTTCCGTTACTACTCAAG 3355  
Qy 1071 AspArg 1072

Db 3356 GACAGA 3361

RESULT 2  
AK084668  
LOCUS  
DEFINITION  
Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched library, clone:D330027H07 product:HYPOTHETICAL 103.9 KDA PROTEIN homolog [Homo sapiens], full insert sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

AK084668 3660 bp mRNA linear HTC 03-APR-2004  
Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched library, clone:D330027H07 product:HYPOTHETICAL 103.9 KDA PROTEIN homolog [Homo sapiens], full insert sequence.

AK084668.1 GI:26351214  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

2 10349636  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

3 20493374  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

4 20530913  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

6 (bases 1 to 3660)  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hasehizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kaga, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, D., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, R., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,

COMMENT  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/.

FEATURES  
source  
1. .3660  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM,DB:D330027H07"  
/db\_xref="taxon:10090"  
/clone="D330027H07"  
/tissue type="heart"  
/dev stage="13 days embryo"  
100.-3357  
/note="unnamed protein product; HYPOTHETICAL 103.9 KDA PROTEIN homolog [Homo sapiens] (SPTRI:AAU55781, evidence: PASTY, 85.5% ID, 97.4% length, match=2763) putative"  
/codon\_start=1  
/protein\_id="BAC39244.1"  
/db\_xref="GI:26351215"  
/translation="MEKGLGKPSASWLVPGYCWTSVKLPRSLYLYSPFCFSVLW LSTDADESRQOQKTLGAGRTGEGHLLRLLAGSLPHACACCRDSACHALWLE GMCQADCSKPSQCPFPPTSSNMLII FQKSTQTTDDIGLLPEDEPHLLGWRGTS WRQSLGAPLTLSPSSHQSLDRQKRLSVVPTGAMQHSKVNSEAGALSPDA SAEVKTITVAGSFTSNHTQTPEWPKNVSITHPESEHSFVPTPOVKSTHSPDTA PLPVAPSYATPTPQASQTSAPHPKELVVSAGKSVQITLTKNVLQNAFLVPE ABPGETVYDMLTHPTDYSGEVGHKHSQSLKSLTPGLYEFVTVDDQNAHGGY VNVVTEPKRPPVAVSPQFQESILPTFTSTIDGQSTDDDKIVQYHWEELGKPL REEKISDITALLKSLVPGNYTSLTVSDGATNSTASLTWNKAVDYPVAVNAGP NOVTLTPNSITLFGNOSTDHHITSYEWSLSPSKGVEMQGVRTPALOLSAMQEG DYTQVLTPTAGQOATQVTVIVQENKPPQADAGPKELTLPVDSVTTLDGSKSTD DQVVSVLNEQSGPDGVLNANSSVATVGLQVGYVTVTLTKDERNLQSSVNV IVKEINPKPVAKIAGNVVTLPTSTAEIDGSRSSDDKIVSYLWTRDETSPAGEVL NNSDHPVFLFLNVEGTITVFLKVTDAKESDTRTTEVVKPDKSNLVEIILDN EVMALTKMLIRIQVGLVGLSDIIVQIQPYTEQSTKMLFVQNDPQLFKGH EVAAKLSKELQKQKADFLIPLALBISTVTCOLGSDHGHCDSTFKRCVCDPFWMENFI KQVLRDGDSCNCEWSVLYVIAFVIVVLAIGLSWTTICCKKROKGRKPKSKYKILDA TDQSLLEKPTSRAGSKQKPTLSLSLHSESELDSDDAITFVTPDREKGLALHWC R SQSLQHLGWWCRAPVCHTSTNSTPGSDSGFCRGGKMLSSDRIYTRITKLSLVH GK"

Alignment Scores:  
Pred. No.: 6,48e-207 Length: 3660  
Score: 2420.50 Matches: 524  
Percent Similarity: 62.16% Conservative: 138  
Best Local Similarity: 49.20% Mismatches: 295  
Query Match: 43.38% Indels: 109  
Gaps: 19  
US-10-764-390-3 (1-1072) x AK084668 (1-3660)

QY 10 SerLeuLeuLeuValThrIleAlaGlyCysAlaArgLysGlnCysSerGluGlyArg 29  
|||||  
Db 220 AGCGTCTGCTGTTGTCACAGATGCTGATGAGAGCAGAG-----TGCCACACAGGGAAG 273  
QY 30 ThrTyrSerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSer 49  
|||||  
Db 274 ACATTTTATGAGCTGGCTGGTGAAGAACTGAGAGGAGAA---AATCATCCTCGGCTTCTTGA 330  
QY 50 HisThrPheProValValAspCysThrAlaAlaCysCysAspLeuSerSerCysAspLeu 69  
|||||  
Db 331 GGAACCTGCTCTTCCACGCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 390  
QY 70 AlaThrPhePheGluGlyArgCysTyrLeuValSerCysProHisLeuGluAsnCysGlu 89  
|||||







[illegible]

Db 1584 CATACTAAACTTAGTAAGCTCGTCCCGGGGAACCTACACTTTCAGCTTAACCTGTTGTGCA 1643

Qy 507 pSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuLeuValAsnAlaValAspTy 527

Db 1644 CTCTGACGGGGTACCAACTCCACACTGCAAGCTGACTGTGAACAAGCTGTGGACTA 1703

Qy 527 rProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleTh 547

Db 1704 CCTCCCGTGGCAATGTCAGGCCCAACCAAGTGTATCACCTGCTCGCTCAGAACTCCATCAC 1763

Qy 547 rLeuAsnGlnAsnGlnSerAspAspHisGlnIleValLeuTyThrGluTrpSerLeuG1 567

Db 1764 CCTCTTTGGAAACCGAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1823

Qy 567 yProGlySerGluGlyLysHisValValMetGlnGlyValGlnThrProTyThrLeuHisLe 587

Db 1824 CCGGAGCACAAGGAAGGTGGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1883

Qy 587 uSerAlaMetGlnGluGlyAspTyThrPheGlnLeuLysValThrAspSerSerArgG1 607

Db 1884 GTCCGCAATGCAAGAGGAGACTATACCTACCAGCTCAGTGTGACACCGCAGGACA 1943

Qy 607 ngInSerThrAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAl 627

Db 1944 ACAGGCCACCGCCCAAGTGAATGTGTGATGTGTGATGTGTGATGTGTGATGTGTGATGTGT 2003

Qy 627 aValAlaGlyProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySe 647

Db 2004 AGATGAGCCCAAGAGAGCTGACCTGCGCGTGGAGACGACACACCTCGAGCGGAG 2063

Qy 647 rSerSerAspAspHisGlyIleValPheTyThrIleThrGluHisValArgGlyProSe 667

Db 2064 CAAGAGCAGATGACAGAGAGTGTCTCTTACCTTTGGAGCAGAGTGGGGACCTGA 2123

Qy 667 rAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValG1 687

Db 2124 CGGGGTGCAGCTGGAGATGCCAACAGACAGTGTGCGCACTGTGCTGGGTGCGCAAGTGG 2183

Qy 687 yThrTyHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrThrLeu 707

Db 2184 GCTTATGTATTACCTTGTACCTGTCAAGATGAGAGAACCTTACAGAGCCAGAGCTTCGT 2243

Qy 707 uThrValAlaValLysLysGluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHi 727

Db 2244 TAATGTCACTGTCAAGAGAGAAATAACAACCGCCAGTAGCCAGATCGCTGGGAACGT 2303

Qy 727 sValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspG1 747

Db 2304 GGTGTGTGACCTTGCCCAAGCAGCAGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTG 2363

Qy 747 nArgIleValSerTyThrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspValI1 767

Db 2364 GGGGATGATGACTACTGTGGACTCGAGATGAGACGCCAGCCGCGAGGGAGGTGT 2423

Qy 767 eAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyThr 787

Db 2424 GAATCACTGTGACCAACCCCTGCTCTCTCTCTCTCAACCTGGTGGAGGGGACCTACAC 2483

Qy 787 rPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValG1 807

Db 2484 GTTTCACCTGAAAGTGACAGATCAAAAGGGCGAGACGACACAGCCGAGCAGCATGTGGA 2543

Qy 807 uValGlnProAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGlyValG1 827

Db 2544 AGTGAAGCTTGACCCCGAGGAAAGCAACCTAGTGGAGATCATCTTGGATGTGAACCTGAC 2603

Qy 827 yGlnLeuThrGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnVa 847

Db 2604 TCAGCTGACTGAGAGGCTGAAGGGATGCTCATCCCGCAGATTTGGGTCTCTCTGGGGT 2663

Qy 847 lLeuAspSerAspIleLysValGlnLysIleArgAlaHisSerAspLeuSerThrValI1 867

Db 2664 GCTGGATTCCGACATCAATTGTGAAAAAGATTTCAGCCGTATACCGGAGCAGACCAAGAT 2723

Qy 867 eValPheTyValGlnSerArgProProPheLysValLeuLysAlaAlaGluValAlaAr 887

Db 2724 GTTGTGTTTGTTCAGAACGACCTCCCCACCAGCTCTTCAAGGCCATGAGTGGCAGC 2783

Qy 887 gAsnLeuHisMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuAr 907

Db 2784 CATGCTCAAGAGCAGCTGCAGAACAGAGAGCTGACTTCTCATCTTTCAGAGCCCTGGA 2843

Qy 907 gValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuTh 927

Db 2844 AATCAGCACATCATGTCTGAGTGAACCTGTTCTGACCATGGCCACTGTGACTCATTTCC 2903

Qy 927 rLysArgCysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyIleTrpAs 947

Db 2904 CAAGCGCTGTCTGTGACCCGTTTGGATGGAGAAATTCATCAAGGTGCAGCTGAGGA 2963

Qy 947 pGlyCysSerAsnCysGluTrpSerIlePheTyThrValThrValLeuAlaPheThrLeuI1 967

Db 2964 TGGAGACAGCAACTGTGAATGGAGCGTCTCTACGTCTCATCTTGTCTTGTCTTGTCTTGT 3023

Qy 967 eValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLysArgTh 987

Db 3024 TGTGCTTGGGATCCTGTCTATGACATCTGCTGCTGCAAGAGCGCAAAA---GG 3080

Qy 987 rLysIleArgLysLysThrLysTyThrIleLeuAspAsnMetAspGluGlnLysArgMe 1007

Db 3081 AAACCCCAAGAGAAAGCAGATACAAGATCCTGGATGCCACAGAT---CAGGAGAGCCT 3137

Qy 1007 tGluLeuArgPro-----LysTyGlyIleLysHisArgSerThrGluHisAsnSerSe 1025

Db 3138 GGAGCTGAACCAACCTCCCGAGCAGGAGCAACAGAAAGGCCCCACCGCTGAGCAGCAG 3197

Qy 1025 rLeuMetValSerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLy 1045

Db 3198 CTTGATGATCTTGAATCGAGCTGGACAGCGAC---GATGCCATCTTTCACATGGCCAGA 3254

Qy 1045 sMetGluArgGly 1049

Db 3255 CCGGAGAGAGGC 3267

RESULT 4

AF289597 3616 bp mRNA linear HTC 01-JAN-2002

LOCUS Homo sapiens clone pp791 unknown mRNA.

DEFINITION AF289597

ACCESSION AF289597

VERSION AF289597.1 GI:18027397

KEYWORDS HTC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Qin,W.X., Zhou,X.M., Zhang,P.P., Jiang,H.Q., Huang,Y., Zhao,X.T., Wan,D.P. and Gu,J.R.

TITLE Novel human cDNA clones with function of inhibiting cancer cell growth

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3616)

AUTHORS Qin,W.X., Zhou,X.M., Zhang,P.P., Jiang,H.Q., Huang,Y., Zhao,X.T., Wan,D.P. and Gu,J.R.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai 200032, P. R. China

FEATURES

source Location/Qualifiers

1..3616

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="pp791"

133..2973

/codon\_start=1

CDS

```
/product="unknown"
/protein_id="AAL55781.1"
/db_xref="GI:18027398"
/translation="MFCQADCSRPSQSCRAFRTHSSNMLVFLKKFQTADDLGLFLPEDD
VPHLLGLGNWASWRQSPRAALRPVSSDQSLIRLQKRSPSDVVTPIVTHQSK
VNDNGLGLTTSGSAEVRKAITISSPLTTDLTAELSGGPKNSVQPEISRLGATTPS
TQOVSKSEKTOIAVQPPVAPSVYATPTPOASFSTSAFYVPIKELVVSAGESVQITL
PKNEVLNAYVLQPPPKGETTYDNQLITHPRDYSGEMGHKSQILKSLKTPGLYEP
KIVQVQNAHGEVNVNTPKEPRKNRPPIAIVSPQFQISLPTITSIDGSQSTDD
KIVTHWHEELGPRLEEKISDITALLKSLVPGNYTFSLTVDSDATNSTTANLV
NKAVDYPVANAGNQVITLPQNSITTLFGNQSTDDHGITSYEWLSPSKGVVEMQG
VTRPTLOLSAMOEGDYTLQVTDITIGQOATQVTVIOPENNNKPPQADAGPKDELTL
PVDSTLQGSKSSDDKIIISYLWEKTOQPGCVOLENANSSVATVGLQVGYVFTLV
KDERNLQSSQVNVIVKEINKPPAKITGNVITLPRSTAEILDGSKSDKGLVSYL
WDRGSAAGGVNHSDDHPLFLSNLVEGTYTFHLKVTDAKESUDRTITVEKDP
KRNKLVEIILIDINVSQITLERLKGFRQIGVLGLVLDSDIIVQIKPYTSQSTKMF
FVQNEPHQIFKGEHVAAMLKSELRKQKADLEIIFRALENTVTCQLNSKHGCHDSFT
KRCCI CDPMFENFTKVQLRDGDSNCMSVLVIATFVIVVALGILSWTVICCKRKQK
GPKRKSKYKTLDTADQESLSLKPTSRAGIKQKGLLLSSLSMHSELSDDAIFTWP
DREKGLLHGQNGSVNPQTPFKARS PREEL"
```

## ORIGIN

## Alignment Scores:

```
Pred. No.: 4,046-204 Length: 3616
Score: 2389.50 Matches: 509
Percent Similarity: 61.26% Conservative: 125
Best Local Similarity: 49.18% Mismatches: 288
Query Match: 42.82% Indels: 113
DB: 3 Gaps: 15
```

US-10-764-390-3 (1-1072) x AF289597 (1-3616)

```
QY 57 CysThrAlaAlaCysCysAspLeuSerSerCysAspLeuAlaTrpTrpPheGluGlyArg 76
DB TGTGGGCTGCTGCTGCAGAGCTCTGCCTGCCATGCTCTTTTGGTGGCTAGAAGGAGT 135
QY 77 CysTyrLeuValSerCysProHisLysGluAsnCysGluProLysLysMetGlyProle 96
DB TGCATTCAGGAGCTGCAGCAGGCCCCAGAGCTGCCGGCTTTTAGGACACACTCTCC 195
QY 97 ArgSerTyrLeuThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyr 116
DB AATTCATGCTGGTGT---TTAAAAAATTCAAACTGCAGATGATTTG----- 243
QY 117 GlyAspMetMetLeuAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAsp 136
DB -----GGCTTTCTACTGAAGAT 261
QY 137 IleArgLysAspLeu***PheLeuGlyLysAspTrpGlyLeuGluGluMetSerGluTyr 156
DB GATGTACCACATCTCTCGGGCTAGGTTCGGAAGTGGCA----- 300
QY 157 **AspAspTyrArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluPro 176
DB -----TCTTGAGGAGGAGGCCA 318
QY 301 ----- 330
QY 177 ArgGlySerAlaGluTyrThrAspTrpGlyLeuLeuProGlySerGluAlaPheAsn 196
DB ----- 333
QY 319 CCCAGAGCTGCACCTC----- 333
QY 197 SerSerValGlyAspSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHis 216
DB -----AGCTTTAATCAGGAAGCTTCAGAAGAGAGGTAGTCCAGT 402
QY 217 TyrLeuAsnGluSerAlaSerThrProAlaProLysLeuProGluArg----- 232
DB ----- 364
QY 233 SerValLeuLeuProLeuProThrProSerSerGlyGluValLeuGluLysGluLys 252
DB -----AAAGTGAATGACTCCACCGAA 456
QY 403 GACGTAGTTACACTATAGTACACAGCATCT----- 456
QY 253 AlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLysGluValLeuMetProSer 272
```

```
DB 457 TTAGGTGCTTGAACACACAGTGGCTTCGACAGAGTCCACAGGCGATTACAAITTCAGT 516
QY 273 HisSerLeuProProAlaSerLeuLeuLeuSer-----SerValThr 286
DB 517 CCCCTAAACACACAGACCTGACTGCAGAGCTGTCTGGTGGGCCAAAAGAAATGTATCAGTCAA 576
QY 287 ValGluLysSerProValLeuThrValThrProGlySer----- 299
DB 577 CCTCAAAATATACAGAGGTCTTGTACTACGCCCCAGCACACAAAGTAAAAAAGTTCTGAG 636
QY 300 ---ThrGluHisSerIleProThrPro-----ProThr 309
DB 637 AAAACCCAGATTGCTGTCCCCCAGCCAGTGGCTCCCTCTCCTACAGTTATGCTACCCCTACC 696
QY 310 SerAlaAlaProSerGluSerThrProSerGluLeuLeuProIleSerProThrAlaPro 329
DB 697 CCCAGGCGCTCTTCCAGAGCACCTCAGCACCATACCAGTT----- 738
QY 330 ArgThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleLeuThrLeuProAsp 349
DB 739 -----ATAAAGGAACCTGGTGTATCTGCTGGAGAGGTGTCAGATAACCTGCTAAG 792
QY 350 AsnGluValGluLeuLysAlaPheValAlaProAlaProProValGluThrThrTyrAsn 369
DB 793 AATCAAGTTCAATTAATGATATGTTCTCCAGAACCCCTTAAAGGAGAACCTACACC 852
QY 370 TyrGluTrpAsnLeuLeuSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHis 389
DB 853 TACGACTGGCAGCTGATTACTCATCTAGACACTACAGTGGAGAAATGGAAGGGGAAACAT 912
QY 390 LysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrVal 409
DB 913 TCCAGATCTCCAAACATATCGAAGCTCTCAGGCTGTATGAATTCAAAGTGAATGTA 972
QY 410 SerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgArg 429
DB 973 GAGGCTCAAAATGCGCCATGGGAAGGCTATGTGAACGTGACAGTCAGCCAGAGCCCGT 1032
QY 430 ValAsnLeuProProValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeu 449
DB 1033 AAGAATCGGCCCCCATTTGCTTATGTGTCTCAGTTCAGTTCAGGAGATCTCTTTGCCAAC 1092
QY 450 ThrSerAlaLeuLeuAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHis 469
DB 1093 ACTTCTCAGTCATTTGATGGCAGTCAAAAGCACTGATGATGATAAATCGTTTCAGTACCAT 1152
QY 470 TrpGluGluIleAsnGlyProPheIleGluLysThrSerValAspSerProValLeu 489
DB 1153 TGGGAAGAACCTTAAGGGGCTCTTAAGAGAGAGAGATTTCTGAAGATACAGCCATATTA 1212
QY 490 ArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAsp 509
DB 1213 AAACCTAAGTAACCTCGTCCCTGGGCACTACACTTTTCAGCTTGTAGTAGACTCTCAT 1272
QY 510 GlyAlaThrAsnSerThrThrAlaAlaLeuIleValAsnAlaValAspTyrProPro 529
DB 1273 GGAGCTACCAACTCTACTACTGCAACCTGACAGTGAACAAAGCTGTGGATTACCCCCCT 1332
QY 530 ValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsn 549
DB 1333 GTGGCCAAACGAGGCCCCAACAAAGTATACCTGCCCCCAAACTCCATCACCTCTCTT 1392
QY 550 GlyAsnGlnSerSerAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGly 569
DB 1393 GGGAAACGAGACCTGATGATCATGGCATCACAGCTATGATGATGATGATGATGATGATGAT 1452
QY 570 SerGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAla 589
DB 1453 AGCAAGGGAAGTGTGGAGATGCGAGGTGTAGAACACCAACCTTACAGCTCTCTCGG 1512
QY 590 MetGlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSer 609
```

Db 1513 ATGCAAGAGGAGACTACACTTACCAGCTCACAGTCACTGACACCAATAGGACAGAGGCC 1572  
 Qy 610 ThrAla\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaValAla 629  
 Db 1573 ACTGCTCAGTACCGCTTATTGTGCACCTGAAACATTAAGCCTCTCTCAGGCAGATGCA 1632  
 Qy 630 GlyProAspGlyeGluLeuPheProValGluSerAlaThrLeuAspGlySerSer 649  
 Db 1633 GGCACCATAAAGAGCTGACCTTCTCTGTGGATAGCACCAACCTGGATGGCAGCAAGAGC 1692  
 Qy 650 SerAspAspHisGlyIleValPheTyrHisTfPTrGluHisValArgGlyProSerAlaVal 669  
 Db 1693 TCAGATGATCAGAAAATTATCTCATATCTCTGGGAAAAAACACAGGACCTGATGGGTG 1752  
 Qy 670 GluMetGluAsnIleAspGlyAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyr 689  
 Db 1753 CAGCTCGGAATGCTAAACAGCAGTGTGTCTACTGTGACTTGGGCTGCAGTGGGGACCTAT 1812  
 Qy 690 HisPheArgLeuThrValIysAspGlnGlnGlyLeuSerSerThrThrLeuThrVal 709  
 Db 1813 GTGTTCACTTCACTGTCAAGATGAGAGGAACCTGCRAAGCCAGAGCTCTGTGAATGTC 1872  
 Qy 710 AlaValIlyGlyeGluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeu 729  
 Db 1873 ATTGTCAAAAGAAATAAACAAACACCTATAGCCAGATAACTGGGAATGTGTGATT 1932  
 Qy 730 ValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIle 749  
 Db 1933 ACCCTACCCAGCAGCACAGAGCTGGATGGCTCTAAGTCTCTCAGATCAAGGAATA 1992  
 Qy 750 ValSerTyrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGly 769  
 Db 1993 GTCAGTACTCTTGGACTCGAGATGAGGGAGCCAGCAGAGGGAGGTGTTAATCAC 2052  
 Qy 770 SerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHis 789  
 Db 2053 TCTGACCATCACCTATCTCTTTCTTCAAACTGGTTGAGGGAACCTTACACTTTTTCAC 2112  
 Qy 790 LeuArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValGlnValGln 809  
 Db 2113 CTGAAGTCAAGGATCAAGGGTGAGGTGACACAGCCGACCACTGTGGAGTGAA 2172  
 Qy 810 ProAspProArgIysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeu 829  
 Db 2173 CTGATCCCAAGAAACAACTGGTGGAGATCATCTTGGATATCACTGATCAGTCA 2232  
 Qy 830 ThrGluGlnArgIysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAsp 849  
 Db 2233 ACTGAGAGGCTGAAGGGATGTTTCATCCGCAATTTGGGTCTCTCTGGGGTGTGGAT 2292  
 Qy 850 SerAspIleIysValGlnIysIleArgAlaHisSerAspLeuSerThrValIleValPhe 869  
 Db 2293 TCGCATCATTTGTGCAAAAGATTTCAGCCGTACACGAGCAGAGCAACCAAAATGTTATT 2352  
 Qy 870 TyrValGlnSerArgProPhePheIysValLeuIysAlaAlaGluValAlaAlaArgLeu 889  
 Db 2353 TTGTGTTCAAAACAGAGCTCCCAACAGATCTTCARAGGCCATGAGTGGCAGCGATGCTC 2412  
 Qy 890 HisMetArgLeuSerIysGluIysAlaAspPheLeuPheIysValLeuArgValAsp 909  
 Db 2413 AAGAGTGAGCTGGGAAGCAAAAGGCAGACTTTTGTATTCAGAGCCTTGGAAAGTCAAC 2472  
 Qy 910 ThrAlaGlyCysLeuLeuIysCysSerGlyHisGlyHisCysAspProLeuThrIysArg 929  
 Db 2473 ACTGTCACATGTCAGCTGAACCTGTTCCGACCATGGCCACTGTGCTGTTCACCAAGCC 2532  
 Qy 930 CysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGlu 949  
 Db 2533 TGATCTGTGACCTTTTGGATGGAGATTTTCATCAAGTGCAGCTGAGGGATGGAGAC 2592  
 Qy 950 SerAsnCysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeu 969  
 Db 2593 AGCAACTGTGAGTGGAGCGTGTATATGTTATCATTTGCTACCTTGTCTATTTGTTGTC 2652

Qy 970 ThrGlyGlyPheThrTrpLeuCysIleCysCysCysLeuArgGlnIysArgThrIysIle 989  
 Db 2653 TTGGGAATCCTGTCTTGGACTGTGATCTGTTGTTGAAGAGCAAAAA--GGAAAAACCC 2709  
 Qy 990 ArgIysIysThrIysTyrThrIleLeuAspAsnMetAspGluGlnGlnArgMetGluLeu 1009  
 Db 2710 AAGAGGAAAGCAAGTACAAGATCTCGATCCACGGAT--CAGGAAAGCTGGAGCTG 2766  
 Qy 1010 ArgPro-----LysTyrGlyIleIysHisArgSerThrGluHisAsnSerSerLeuMet 1027  
 Db 2767 AAGCAACCTCCAGCAGGAGCATCAACAGAAAGGCCCTTTTGCTAAGTAGGACGCTGATG 2826  
 Qy 1028 ValSerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluIysMetGlu 1047  
 Db 2827 CACTCCGAGTCAGAGCTGCACAGCGAC---GATGCCATCTTTACATGCCAGACCGAGAG 2883  
 Qy 1048 ArgGlyAsnProIysValSerMetAsnGlySerIleArgAsnGly 1062  
 Db 2884 AAGGCAAACTCCTGTCATGGTGCAGATGGCTCTGTACCAACGGG 2928  
 RESULT 5  
 AK031342 3067 bp mRNA linear HTC 03-APR-2004  
 LOCUS Mus musculus 13 days embryo male testis cDNA, RIKEN full-length  
 DEFINITION enriched library, clone:6030408013 product:HYPOTHETICAL 103.9 KDA  
 PROTEIN homolog [Homo sapiens], full insert sequence.  
 ACCESSION AK031342 GI:26082279  
 VERSION AK031342.1  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 1 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multiplexed sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 6 (bases 1 to 3067)



1398	Db	CCCCGTAAGAACCGCGCTCCGGTTGCTGGTGTCTACCTCAGTTCCAGAGAGATCTCGCT	1457
447	Qy	uProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspSerThrGluIleValse	467
1458	Db	GCCAAACCACTTCTACCATCATTTGATGCGAGCCAGAGCAAGATGATATAAATTTGTCCA	1517
467	Qy	rTyHisTrrpGluGluIleAsnGlyProPheIleGluGluIysThrSerValAspSerPr	487
1518	Db	GTACCACTGGGAAGAGCTTAAGGGGCGCCCTGAGAGAAGAAGATCTCTCTGAACACACAGC	1577
487	Qy	oValLeuArgLeuSerAsnLeuAspProGlyAsnTySerPheArgLeuThrValThrAs	507
1578	Db	CATACTAAACCTTAGTAGCTGTCGCCGGGAATACACCTTCAGCTTAACCTGTGTGCGA	1637
507	Qy	pSerAspGlyAlaThrAsnSerThrThrAlaLeuIleValAsnAsnAlaValAspTy	527
1638	Db	CTCTGA CGGGGTACCAACTCCACCACTGCAGAGCTGACTGTGNACAAAGACTGTGACTA	1697
527	Qy	rProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleTh	547
1698	Db	CCCTCCGCTGGCCAAATGAGCGCCCAACCAAGTGATCACCTGCTCAGAACTCCATCAC	1757
547	Qy	rLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuTyrrGluTrrpSerLeuGl	567
1758	Db	CCTTTTGGAAAC CAGAGCA CGGATGACCA CGGCGATCACCA GCTATGAGTGGTGGCTCAG	1817
567	Qy	yProGlySerGluGlyIysHisValValMetGlnGlyValGlnThrProTyrrLeuHisLe	587
1818	Db	CCCGAGACCAAGGGGAAGGTGGTGGAGATGTCAGGGAGTTAGAACGCCAGCCCTGCAGCT	1877
587	Qy	uSerAlaMetGlnGluGlyAspTyrrThrPheGlnLeuLysValThrAspSerSerArgGl	607
1878	Db	GTCCGCAATGCAAGAGAGACTATACCTACCACTCACAGCTCACACCGCAGGACACA	1937
607	Qy	nGlnSerThrAla**ValThrValIleValGlnProGluAsnAsnArgProProValAl	627
1938	Db	ACAGGCCACCCCAAGTGACTGTGATTGTGTGCAGCTTGAGAACAAACAAGCTCTCTCAGGC	1997
627	Qy	aValAlaGlyProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySe	647
1998	Db	AGATGCAGGCCCAACAAGAGCTGACCCCTGCCGTGGCAGACACAACCTTGACGGCAG	2057
647	Qy	rSerSerSerAspAspHisGlyIleValPheTyrrHisTrrpGluHisValArgGlyProSe	667
2058	Db	CAAGGACACAGATGCCAGAGAGTCTCTTACTTTGGGAGCAGAGTCCGGGACCTGA	2117
667	Qy	rAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGl	687
2118	Db	CGGGGTGCAGCTGGAGAATGCCAACGACGAGTGTGCCACTGTACTGGGCTGCAAGTCGG	2177
687	Qy	yThrTyrrHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLe	707
2178	Db	GACTTATGTATTCACCTTGACTGTCAAGATGAGAGGAACCTACAGAGCCAGAGCTCCGT	2237
707	Qy	uThrValAlaValLysLysGluAsnAsnSerProProArgAlaGlyGlyArgHis	727
2238	Db	TAATGTCTATTGTCAAGAGAGAAATAAACAACCCCGTAGTACCAAGATCGCTGGGAACGT	2297
727	Qy	sValLeuValLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGl	747
2298	Db	GGTGGTGCACCTTGCCACGACACAGCTGAGCTGGATGGCTCGAGGTCTCCAGATGACAA	2357
747	Qy	nArgIleValSerTyrrLeuTrrIleArgAspGlyGlnSerProAlaIleGlyAspValI	767
2358	Db	GGGGATAGTTCAGCTACTGTGTGACTCGAGATGAGACGAGCCACCGCAGGGAGGTGCT	2417
767	Qy	eAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrrTh	787
2418	Db	GAATCACTCTGACCAACCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2477
787	Qy	rPheHisLeuArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValGl	807
2478	Db	GTTCACCTGAAGATGACAGATGCAAGGGCGAGGAGGACACACACCGACGACAGTGGGA	2537

Qy	807	uValGlnProAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGlyValGI	827
Db	2538	AGTGAAGCCTGACCCAGGAAAGCAACCTAGTGGAGATCATCTTGGATGTGAACGTCAG	2597
Qy	827	yGlnLeuThrGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuAsnVa	847
Db	2598	TCAGCTGACGTGAGAGCTGAAGGGAGTGTCTATCCCGCCAGATTTGGGGTCTCTCTGGGGGT	2657
Qy	847	lLeuAspSerAspIleLysValGlnLysIleArgAlaHisSerAspLeuSerThrValII	867
Db	2658	GCTGGATTCGACATCATTTGTGCAAAAGATTACGCGGTACACGGAGCAGACCAAGAT	2717
Qy	867	eValPheTyrValGlnSerArgProProPheLysValLeuLysAlaGluValAlaAr	887
Db	2718	GTGTGTTTTTTGTTTCAGAACGACCTCCCAACACAGCTCTTCAAAGGCCATCAGGTGGCAGC	2777
Qy	887	gAsnLeuHisMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuAr	907
Db	2778	CATGCTCAAGAGCGAGCTGCAGAACGAAAGGCTGACTTCTCTCATCTTCAGAGCCCTGGGA	2837
Qy	907	gValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuTh	927
Db	2838	AATCAGCACAGTCACATGTCAGCTGAACTGTTCTGACCATGGCCATGTGACTCATCTCAC	2897
Qy	927	rLysArgCysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAs	947
Db	2898	CAAGCGCTGTCTGTGACCCGTTTTGGATGGAGAATTTTCATCAAGTGCAGCTGAGGGA	2957
Qy	947	pGlyGluSerAsnCysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuII	967
Db	2958	TGGAGACGACCACTGTAATGGAGCGTGTCTACGTTCATCATTCGCTCTCTTGTCTCATGT	3017
Qy	967	eValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysCysLeuArg	983
Db	3018	TGTTGCTTGGGGATCTGTGTCATGGACTACAACTGCTGCTGTCACAGAGG	3066
RESULT 6			
HSM805357			
LOCUS			
DEFINITION	HSM805357	3671 bp	mRNA linear
ACCESSION	Homo sapiens mRNA; cDNA DKFZp434C0829 (from clone DKFZp434C0829).		
VERSION	AL834315		
KEYWORDS	AL834315.1	GI:21739924	
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Koehler, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo G., Han, M. and Wiemann, S.		
CONSRTH	The German cDNA Consortium		
TITLE	Direct Submission		
JOURNAL	Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764		
COMMENT	Neuherberg, Germany Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMFZ (Biomedical Research Center at the Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp434C0829) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp434C0829 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.		
FEATURES	Location/Qualifiers		
SOURCE	1..3671		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="RZPD:DKFZp434C0829"		
	/db_xref="taxon:9606"		
	/clone="DKFZp434C0829"		





```

Db 1568 ATCATTTGCAAAAGATTACAGCGGTACAGGAGCAGACCAAAATGGTATTTTGT 1627
QY 872 GlnSerArgProPhelysValLeuLysAlaIaGluValAlaAraAsnLeuHisMet 891
Db 1628 CAAACAGAGCTTCCCAACATCTTCAAGGCCATGAGGTGGCAGCATGCTCAAGAGT 1687
QY 892 ArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspThrAla 911
Db 1688 GAGCTGCGGAGCAAAAGGACAGACTTTTGATTTACAGAGCCTTGGAAAGTCAACTGTC 1747
QY 912 GlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCysIle 931
Db 1748 ACATGTCAGTGAAGCTTCCAGACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1807
QY 932 CysSerHisLeuTrpMetGluAsnLeuLysGlnArgTyrIleTrpAspGlyGluSerAsn 951
Db 1808 TGTGACCTTTTGGATGGAGAAATTCATCAAGGTGCAGCTGAGGATGGAGACGAAC 1867
QY 952 CysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuThrGly 971
Db 1868 TGTGAGTGAGCGGTGTAATGTTATCATGCTACTTGTGCTGCTGCTGCTGCTGCTG 1927
QY 972 GlyPheThrTrpLeuLysCysCysCysLysArgGlnLysArgThrLysIleArgLys 991
Db 1928 ATCTGCTGTGACTGTGATCTGTTGTTGTAAGAGGCAAAA--GGAAAAACCAAGAGG 1984
QY 992 LysThrLysTyrThrIleLeuAspAsnMetAspGluGlnGlnLysArgMetGluLeuArgPro 1011
Db 1985 AAAAGCAAGTACAAGATCTGATGACCAAGGAT--CAGGAAGCTCTGAGTGAAGCCA 2041
QY 1012 -----LysTyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSer 1029
Db 2042 ACCTCCGAGCGAGGATCAACAGAAAGCCCTTTGCTAAGTAGCAGCCTGATGACTCC 2101
QY 1030 GluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGly 1049
Db 2102 GAGTCAGAGCTGACAGCGAT--GATGCCATCTTTACATGCCAGCAGCAGAGAGGCG 2158
QY 1050 AsnProLysValSerMetAsnGlySerIleArgAsnGly 1062
Db 2159 AAACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2197

```

```

RESULT 7
AY417695 2959 bp DNA linear GSS 12-DEC-2003
LOCUS Mus musculus HCMG309 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY417695
VERSION AY417695.1 GI:39773655
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2959)
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferred nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
JOURNAL 14671302
PUBMED 2 (bases 1 to 2959)
REFERENCE Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
AUTHORS Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering

```

```

FEATURES
source 1..2959
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>2959
/locus_tag="HCM6309"
ORIGIN
Alignment Scores:
Pred. No.: 2,06e-167 Length: 2959
Score: 1982.00 Matches: 448
Percent Similarity: 55.32% Conservative: 119
Best Local Similarity: 43.71% Mismatches: 353
Query Match: 35.52% Indels: 106
DB: 9 Gaps: 17
US-10-764-390-3 (1-1072) x AY417695 (1-2959)
QY 10 SerLeuLeuLeuValThrIleAlaGlyCysAlaArgLysGlnCysSerGluGlyArg 29
Db 121 AGCGTTCTGTGGTGTCAACAGATGCTGATGAGAGCAGA-----TGCCAACAGGGGAAG 174
QY 30 ThrTyrSerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSer 49
Db 175 ACATTTATGAGCTGGCTTGAGAACTCAGGGAGAA---AATCACCTCCGGCTTCTTGA 231
QY 50 HisThrPheProValValAspCysThrAlaAlaCysCysAspLeuSerSerCysAspLeu 69
Db 232 GGAAGCCTGCTTCCACAGCCTGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
QY 70 AlaTrpTrpPheGluGlyArgCysTyrLeuValSerCysProHisLysGluAsnCysGlu 89
Db 292 CTATGGTGGCTGGAAGGATGTGCTTTCAGGCTGACTGACTGACTGACTGACTGACTG 351
QY 90 ProLysLysMetGlyProIleArgSerTyrLeuThrPheValLeuArgProValGlnArg 109
Db 352 CCTTTAGACAGACTCTTCCAATTCATGCTG----- 384
QY 110 ProAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgGlySerProSerGlyIle 129
Db 385 -----ATCATTTTCAAAAAATCCAAACTACAGATGAT 417
QY 130 TrpGlyAspSerProGluAspIleArgLysAspLeu***PheLeuGlyLysAspTrpGly 149
Db 418 TTGGGCTTCTGCTGGAAGATGATGAACCATCTTCTGAGCTAGGC-----TGGGCG 471
QY 150 LeuGluGluMetSerGluTyr***AspAspTyrArgGluLeuGluLysAspLeuGln 169
Db 472 AG-GACATCGTGGAGGAG----- 494
QY 170 ProSer---GlyLysGlnGluPro-----Arg 177
Db 495 CCTTCTTGGGCTCCCTCCCTTCTGTATCCCTCTAGTCACCCAGAGAGCTTACTCAG 554
QY 178 GlySerAla---GluTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPhe--- 195
Db 555 GGATCGGAGAGAGAGATCTCAGTGTGTTACCTACATGAGGAGGAGTGCAGCATTTCTAA 614
QY 196 AsnSerSerValGlyAspSerProAlaValProAlaGluThrGlnGlnAspProGluLeu 215
Db 615 AGTGAATCACTCGGAGGAGGAGGTGCTCTGAGTCCCACTCTGAGAGGTCGGCAAAAC 674
QY 216 Histyr-----LeuAsnGlu-SerAlaSerThrProAlaProLysLeuProGlu 231
Db 675 CATTACAGTGGCGGTCTCTTCCACCAAGTAACCACTACACAGACTCTCTGAGTGGCCAA 734
QY 231 uArgSerValLeuLeuProLeuLeuThrThrProSerSerGlyGluValLeuGluLysGlu 251
Db 735 GAATGTGCTCCATCCATCTGAACTCC----- 762
QY 251 uLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLysGluValLeuMetPr 271

```





```

Qy 414 aPheGlyGluGlyPheValAsnValThrValIysProAlaArgValAsnLeuProPr 434
Db 1164 CCATGGGGAAGGCTATGTGAACCTGACAGTCAAGCCAGAGCCCGCTAAGAATCGGCCCC 1223
Qy 434 oValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeu1 454
Db 1224 CATGTCTATTGTGTCACCTCAGTCCAGAGATCTCTTTGCCAACCACTTCTCAGCTCAT 1283
Qy 454 eAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTrpGluIleAs 474
Db 1284 TGATGGCAGTCAAGCAGCTGATGATGATAAAATCGTTCAGTACCATTTGGGAAGACTTAA 1343
Qy 474 nGlyProPheIleGluGlyThrSerValAspSerProValLeuArgLeuSerAsnLe 494
Db 1344 GGGCCCTCTAAGAGAAGAGATTTCTGAAGATACAGCCATATATTAACACTAAGTAAACT 1403
Qy 494 uAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSe 514
Db 1404 CGTCCCTGGGAACCTACCTTTACGCTTGATGTAGTACTCTGATGAGCTACCACTC 1463
Qy 514 rThrThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProProValAlaAsnAlaGl 534
Db 1464 TACTACTGCAAACTGACAGTGAACAAAGCTGTGGATTACCCCTGTGGCCAAAGCAGG 1523
Qy 534 yProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSe 554
Db 1524 CCCCAACCAAGTGTACCTCGCCCAAACTCCATCACTCTTTGGGAACCAAGAGCAC 1583
Qy 554 rAspAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyIleH 574
Db 1584 TGATGATCATGGCATCACCACTATGATGTGCTACTCAGCCCAAGCAAGAGGAAGT 1643
Qy 574 sValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGlyIleAs 594
Db 1644 GTTGAGATGAGGGTGTAGAACACCACTTACAGCTCTCTCGGATGCAAGAGGAGA 1703
Qy 594 pTyrThrPheGlnLeuIleValThrAspSerArgGlnGlnSerThrAla***ValTh 614
Db 1704 CTACACTTACCAGCTCAGCTGACTGACACAAAGGACAGCAGGCGCACTGCTCAAGTGAC 1763
Qy 614 rValIleValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLysGl 634
Db 1764 TGTATTGTGCAACTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1823
Qy 634 uLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerSerAspHisGl 654
Db 1824 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1883
Qy 654 yIleValPheTyrHisTrpGluHisValArgGlyProSerAlaValGluMetGluAsnIl 674
Db 1884 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1943
Qy 674 eAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuTh 694
Db 1944 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2003
Qy 694 rValLysAspGlnGlnGlyLeuSerSerThrThrLeuThrValAlaValLysLysGl 714
Db 2004 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2063
Qy 714 uAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAs 734
Db 2064 AATAACAAACCACTATAGCAAGATACTGGGAATGTGTGATTAACCTTACCACCCAGAG 2123
Qy 734 nSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuTr 754
Db 2124 CACAGCAGAGTGTGCTCTAAGTCTCAGATGACAAAGGGAATGATGACAGTACCTCTG 2183
Qy 754 pIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerVa 774
Db 2184 GACTCGAGATGAGGGGAGCCCGCAGCAGCAGGAGGAGGTGTAAATCACTCTGACCATCACCC 2243

```

```

Qy 774 lAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrAs 794
Db 2244 TATCTTTTCTTTTCAAAACCTGGTTGAGGGAACCTACACTTTTCACTTGAAGTGACCGA 2303
Qy 794 pSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArgLys 814
Db 2304 TGCAAAAGGTTGAGAGTGCACAGACCGGACCACTGTGGAGGTGAACCTGATCCAGAGA 2363
Qy 814 sSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLys 834
Db 2364 AAACAACCTGGTGAGATCATCTTGGATATCAACGTCAGTCAGCTAACTGAGAGGCTGAA 2423
Qy 834 sAspThrLeuValArgGlnLeuAlaValLeuAsnValLeuAspSerAspIleLysVa 854
Db 2424 GGGGATGTTCACTGCCAGATTTGGGTCCTCTGGGGTGTGTGATTCGACATCATTTGT 2483
Qy 854 lGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSerAr 874
Db 2484 GCAAAAGATTGAGCGGTACAGGAGCAGAGCAACAAATGATATTTTGTTCAAACGA 2543
Qy 874 gProProPheLysValLeuLysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSe 894
Db 2544 GCCTCCCCACCAGATCTTCAAGGCCCATGAGTGGCAGCGATGCTCAAGAGTGAGCTGCG 2603
Qy 894 rLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCysLe 914
Db 2604 GAAGCAAAAGCGACAGATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2663
Qy 914 uLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHi 934
Db 2664 GCTGACTGTTCCGACCATGGCCACTGTGATCTGTTTACCAACGCTGATCTGTGACCC 2723
Qy 934 sLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGluSerAsnCysGluTr 954
Db 2724 TTTTGTGATGAGAAATTTTCATCAAGTGCAGCTGAGGATGGAGACAGCACTGTGAGTG 2783
Qy 954 pSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheTh 974
Db 2784 GAGCGCTTATATTTATCTATCTTGTACCTTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 2843
Qy 974 rTrpLeuCysLysCysCysLysArgGlnLysArgThrLysIleArgLysLysThrLys 994
Db 2844 TTGACTGTGATCTGTTGTTGTAAGAGGCAAAA---GGAAACCCAGAGAGAAAGCA 2900
Qy 994 sTyrThrIleLeuAspAsnMetAspGluGlnGlnArgMetGluLeuArgPro 1011
Db 2901 GTACAAGATCTGGATGCCACGGAT---CAGAAAGCTGGAGCTGAAGCCA 2949

RESULT 9
AY417694
LOCUS
DEFINITION Pan troglodytes HCM6309 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY417694
VERSION AY417694.1 GI:39773654
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 2962)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2962)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,

```

Adams,M.D. and Cargill,M.  
Direct Submission  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.

FEATURES  
source Location/Qualifiers  
1..2962  
    /organism="Pan troglodytes"  
    /mol\_type="genomic DNA"  
    /db\_xref="taxon:9598"  
    <1\_>2962  
    /locus\_tag="HCM6309"

gene

ORIGIN

Alignment Scores:			
Pred. No.:	4,96e-136	Length:	2962
Score:	1635.00	Matches:	386
Percent Similarity:	52.56%	Conservative:	107
Best Local Similarity:	41.15%	Mismatches:	374
Query Match:	29.30%	Indels:	72
DB:	9	Gaps:	13

US-10-764-390-3 (1-1072) x AY417694 (1-2962)

```

Qy      82   CyeProHisGlycLuanCyeGluProLysLysMetGlyProLleArzSerTyLeuThr 101
        |||          :|||:|||||          :|||          |||
Db     328   TGCAGCAGGCCCCAGACGTGCGGGCTTTTAGACACACACTCTCCAATTCATCGTGGTG 387
        |||          :|||:|||||          :|||          |||
Qy     102   PheValLeuArgProValGlnArgProAlaGlnLeuLeuasPtyrGlyAspMetLeu 121
        |||          :|||:|||||          :|||          |||
Db     388   TTT---TTAAAAAATTCCAACTCGCATGATTG----- 420
        |||          :|||:|||||          :|||          |||
Qy     122   AsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArzLysAspLeu 141
        |||          :|||:|||||          :|||          |||
Db     421   -----GGCTTTCTACCTGAAGATGATGTACCACATCTT 453
        |||          :|||:|||||          :|||          |||
Qy     142   ***PheLeuGlyLysAspTrpGlyLeuGluMetSerGluTy***AspAsp-TyrAr 161
        |||          :|||:|||||          :|||          |||
Db     454   CTGGGGCTAGGTGGAACTGG-----GCATCTTGGAGCGAGACCCCACCC 498
        |||          :|||:|||||          :|||          |||
Qy     161   gGluLeuGluLysAspLeuGlnPro-----SerGlyLysGI 174
        |||          :|||:|||||          :|||          |||
Db     499   AGAGCTGCACCTAGACCTGCTGTATCTCCAGTCAGCAGACAGCTTAATCAGAAGACTT 558
        |||          :|||:|||||          :|||          |||
Qy     174   nGluProArzGlySerAlaGluTyThrAspTrpGlyLeuLeuProGlySerGluGlyAl 194
        |||          :|||:|||||          :|||          |||
Db     559   -CAGAASAGAGAGGTAGTCCCGAGTGAAGTAGTTACACCTATAGTG-----AC 602
        |||          :|||:|||||          :|||          |||
Qy     194   aPheAnSerSerValGlyAspSerProAlaValProAlaGluThrGlnGlnAspProGl 214
        |||          :|||:|||||          :|||          |||
Db     603   ACAGCACTTCAAAGTGAATGACTCC-----ANNGA 632
        |||          :|||:|||||          :|||          |||
Qy     214   uLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeuProGluArgSerVa 234
        |||          :|||:|||||          :|||          |||
Db     633   ATTAGNNNGTCTGACTACCACTGGCGTCTGCAGAGGTCCACAAG-----GGGAT 680
        |||          :|||:|||||          :|||          |||
Qy     234   lLeuLeuProLeuProThrThrProSerSerGlyGluValLeuGluLysGluLysAlaSe 254
        |||          :|||:|||||          :|||          |||
Db     681   TACAATTTCCAGTCCCTAACCC-----ACAGACCTGACTGC 716
        |||          :|||:|||||          :|||          |||
Qy     254   xGlnLeuGlnGluInSerSerAnSerSerGlyLysGluValLeuMetProSerHisSe 274
        |||          :|||:|||||          :|||          |||
Db     717   AGAGCTGTCGTGGTGCGCCAAAAGATGTATCAGTCGCACTCGAAATATCAGAGGGTCTTGC 776
        |||          :|||:|||||          :|||          |||
Qy     274   rLeuProProAlaSerLeuGluLeuSerSerValThrValGluLysSerProValLeuTh 294
        |||          :|||:|||||          :|||          |||
Db     777   TACTACGCCCAGCACCTACACAGTAAAGT-----TCTGAGAAACCCAGATGCTGCT 830
        |||          :|||:|||||          :|||          |||
Qy     294   rValThrProGlySerThrGluHisSerIleProThrProThrSerlAlaProse 314
        |||          :|||:|||||          :|||          |||
Db     831   CCCCCCAGCAGTGGCTCCCTCCTACAGTTATGTACTACC---CCTACCCCCCANGCANNNTT 887
        |||          :|||:|||||          :|||          |||
    
```





URL:http://fantom.gsc.riken.jp/. Location/Qualifiers		150 LeuGluGluMetSerGluTyr***AspAspTyrArgGluLeuGluLysAspLeuLeuGln 169	
FEATURES	Source	1. .2189	609 AG-GACATCGTGGAGGAG-----GCAGAG 631
/organism="Mus musculus"		170 ProSer---GlyLysGlnGluPro-----Arg 177	
/mol_type="mRNA"		632 CCTTCTTGGGGCTCCCTCACCTTCTGTACCTCTAGTACCACACAGAGCTTACTCAG 691	
/strain="CS7BL/6J"		178 GlySerAla---GluTyrThrAspTyrGlyLeuLeuProGlySerGluGlyAlaPhe--- 195	
/db_xref="FANTOM_DB:C430042D23"		692 GGATCGGACAGAGAGATCTCAGTGTGTACCTACATGAGGAGGATGTCAGATTCTTAA 751	
/db_xref="taxon:10090"		196 AsnSerSerValGlyAspSerProAlaValProAlaGluThrGlnGlnAspProGluLeu 215	
/clone="C430042D23"		752 AGTAAATCACTCGGAGGAGCAGGTGTCTGTAGTCCACCTCTCGCAGAGGTCGCGAAAC 811	
/tissue_type="whole body"		216 HisTyr-----LeuAsnGlu-SerAlaSerThrProAlaProLysLeuProGln 231	
/clone_lib="RIKEN full-length enriched mouse cDNA library"		812 CATTACAGTTGCCGGTTCCTTCCAGTAACACACATACACAGACTCTGTAGTGGCCCA 871	
/dev_stage="7 days embryo"		231 uArgSerValLeuLeuProLeuProThrThrProSerSerGlyGluValLeuGluLysGln 251	
138. .>2189		872 GAATGTGTCCATCTCCTGAACCATCC----- 899	
/note="unnamed protein product; HYPOTHETICAL 103.9 KDA PROTEIN homolog [Homo sapiens] (SPTR AAL55781, evidence: FASTV, 85.5%ID, 97.4%length, match=2763) putative"		251 uLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyLysGluValLeuMetPr 271	
/codon_start=1		900 -----GAGCAGCTCAGTCTGTATCTGGT----- 923	
/db_xref="GI:26340310"		271 oSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGluLysSerPr 291	
/translation="MEKRLGVKPSPAWVLPYCWOTSVKLPRLSLYLILYSPFCFSLW LSTDADESRCCQGLYAGLRTEGENHLRLAGSLPHACRAACRDSACHALWMLE GMCFOADSKQSCOPPTDSDNSMLIIFQKSQTDDLGILLPEDDEPHLLRLGWRIS WRKQLGAPILFVPSHHQSLDRDRDLSSVPTGAMQHSKVNHSERAGALSP SAEVRKTTIVAGSFTSNHTTQPEMPKNVSHPESEHSSPSVSGTPQVKSHEGPTDA PLVPASYSYATPTQASQSSTAPHPVKLSAGKSVQITLPKNEVLQNALFVLE AEPGTYSYDGLITHTPDYSGEVERKHSQSLQSLKTPGLYEFKVTVDQNAHGEY VNVTVKPRKRPVAVVPQFQISLPTSTIIDGSSQSTDDDKIVQYHWEELKGPL REEKISBETAILKLSLVPNGYTFSLTVDSGATNSTASLTYNKAVDYPVNAVGP NOVITLPSNTLFGNSDTHGITSYSPSKGVKEMQGVRTPALQSLAMQBG DYTYQLTDTAGQAQAVTVIQPENKPPQADAGPKELTLPVDSSTTLDGSKSTD DQVVSYLWEQSRGPDGVQLENANSSVATVTVLQGTIVFTLVKDERNLQSSVNV 1A"		923 ----- 923	
CDS		291 oValLeuThrValThrProGlySerThrGluHisSerIleProThrProThrSerAl 311	
ORIGIN		924 -----ACTCCGCAAGTAAAGACACTGAGCAGTCAACTGATGCCCTCTGCCAGT 976	
Alignment Scores:		311 aAlaProSerGluSer-----ThrProSer---GluLeuProIleSerProThrThrAl 328	
Pred. No.:	2,048-121	Length:	2189
Score:	1471.00	Matches:	337
Percent Similarity:	58.23%	Conservative:	84
Best Local Similarity:	46.61%	Mismatches:	199
Query Match:	26.36%	Indels:	104
DB:	3	Gaps:	15
US-10-764-390-3 (1-1072) x AK049570 (1-2189)		1037 ACCACACCCAGTTGTAAAGGAGCTGGTGTCTGTCTGGGAGAGCGTCCAGATCACCCT 1096	
Qy	10 SerLeuLeuLeuValThrIleAlaGlyCysAlaAatGlyGlnCysSerGluGlyArg 29	Qy	347 uProAspAsnGluValGluLeuLysAlaPheValAlaProAlaProValGluThrTh 367
Db	258 AGCGTTCTGTGTGTCAACAGATGCTGATGAGACAGA-----TGCCAACAGGGGAAG 311	Db	1097 GCCTAAGNATGAAGTTCAGTTAATGCTTCCTTCCAGACAGAGCCAGGAGAAC 1156
Qy	30 ThrTyrSerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSer 49	Qy	367 rTyrAsnTyrGluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluIleLysGln 387
Db	312 ACACTTTATGGAGCTGGCTTGAGACTGAGGGAGAA---AATCACCTCCGGCTTCTTGA 368	Db	1157 CTACACCTACGACTGGCAGCTGATCACTCATCTACAGACTACAGTGTGAGAGGTTGAGAG 1216
Qy	50 HisThrPheProValValAspCysThrAlaAlaCysCysAspLeuSerCysAspLeu 69	Qy	387 nGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysVa 407
Db	369 GGAAGCTCCCTTTCACGCCCTGTGGGCTGCTGCTGCGGAGACTCTGCTGCCACGCT 428	Db	1217 GAAACATTCACAGAGCTCCAACTGTCCAACTGCTCAAGCTCAGCCCTGTAGAAATTCAGGT 1276
Qy	70 AlaTrpTrpPheGluGlyArgCysTyrLeuValSerCysProHisLysGluAsnCysGlu 89	Qy	407 lThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAl 427
Db	429 CTATGTGTGCTGGAAGGATGCTTTTCAGGCTGACTGCAGTAAGCCCCAGAGCTGCCAG 488	Db	1277 GACTGTGATGGCCAGATGCTCCATGGGAGGCTACGTGAATGTGACAGTGAACACAGA 1336
Qy	90 ProLysLeuMetGlyProIleArgSerTyrLeuThrPheValLeuArgProValGlnArg 109	Qy	427 aArgArgValAsnLeuProProValAlaValSerProGlnLeuGlnGluLeuThrIle 447
Db	489 CTTTATGACACAGACTCTTCCATTCATGCTG----- 521	Db	1337 GCCCGTAAAGAACCGGCTCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1396
Qy	110 ProAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgLysProSerGlyIle 129	Qy	447 uProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSe 467
Db	522 -----ATCAUTTTTCAAAAAATCCCAAACTACAGATGAT 554	Db	1397 GCCAACCACTTCTACCATCATTTGATGGCAGCCAGCAGCAGATGACGATAAATTTGTCCA 1456
Qy	130 TrpGlyAspSerProGluAspIleArgLysAspLeu***PheIleuGlyLysAspTrpGly 149	Qy	467 rTyrHisTrpGluGluLeuAsnGlyProPheIleGluLysThrSerValAspSerPr 487
Db	555 TTGGGCGCTTCTGCTGAAGATGATGAACCAACCATCTTCTGAGGCTAGGC-----TGGGCG 608	Db	1457 GTACACCTGGGAGAGCTTAAGGGGCCCCCTGAGAGAGAGAGATCTCTGAGAGACACAGC 1516
Qy		Qy	487 oValLeuArgLeuSerAsnLeuLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAs 507

```
Db 1517 CATACTAAACTTAGTAACTGCTCCCGGGAACCTACACCTTACAGCTTAAGCTGTGTGCGA 1576
Qy 507 pSerAspGlyValThrAsnSerThrAlaAlaLeuLeuValAsnAsnAlaValAspTy 527
Db 1577 CTCTGACGGGGCTACCAACTCCACCACTGCAAGCTGACTGTGAACAAGCTGTGGACTA 1636
Qy 527 rProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleTh 547
Db 1637 CCCTCCCGTGGCCCAATGCGAGCCCAACCAAGTGTATCACCTGTGCTCAGAACTCCATCAC 1696
Qy 547 rLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuTyTrpSerLeuG1 567
Db 1697 CCTCTTTGGAAACACAGACGACGAGATGACCGGATACACGCTATGAGTGTGCTCAG 1756
Qy 567 yProGlySerGluGlyHisValValMetGlnGlyValGlnThrProTyTrpLeuHisLe 587
Db 1757 CCCGAGCAGCAAGGGAAGTGGTGGAGATGACGGAGTTAGAACGCCAGCCCTGCAGCT 1816
Qy 587 uSerAlaMetGlnGluGlyAspTyTrpPheGlnLeuLysValThrAspSerSerArgL 607
Db 1817 GTCCGCAATGCAAGAAGGAGACTATACCTACAGCTCACAGTGTGACACCGGAGACA 1876
Qy 607 nGlnSerThrAla**ValThrValIleValGlnProGluAsnAsnArgProProValAl 627
Db 1877 ACAGCCACCCCAAGTACTGTGATGTGACCGCTGGAACACAGCCCTCTCAGGC 1936
Qy 627 aValAlaGlyProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySe 647
Db 1937 AGATGACGGCCACACAAAGAGCTGACCTGCCGTGGACAGCAACCCCTGGACGGCAG 1996
Qy 647 rSerSerSerAspAspHisGlyIleValPheTyTrpHisThrGluHisValArgGlyProSe 667
Db 1997 CAAGAGCAGACAGACAGAGAGTGTCTTACCTTTGGAGCAGAGTGTGGGACCTGA 2056
Qy 667 rAlaValGluMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValG1 687
Db 2057 CGGGGTGACGTGGAGAAATGCCAAGCAGGTGTGCCACTGTGACTGTGTGCAAGTCGG 2116
Qy 687 yThrTyHisPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrThrLe 707
Db 2117 GACTTATGATTCACCTTGACTGTCTCAAGATGAGAGGACCTACAGAGCCAGAGCTCGT 2176
Qy 707 uThrVal 709
Db 2177 TAATGTC 2183

RESULT 11
AUI30115 LOCUS AUI30115 842 bp mRNA linear EST 01-AUG-2002
DEFINITION AUI30115 NT2RP3 Homo sapiens cDNA clone NT2RP3000279 5', mRNA
sequence.
ACCESSION AUI30115
VERSION AUI30115.1 GI:10990469
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 842)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
```

```
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
Source
1. 842
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP3000279"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RP3"
/notes="Vector: pME18SF13; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

ORIGIN
Alignment Scores:
Pred. No.: 1,01e-107 Length: 842
Score: 1312.00 Matches: 267
Percent Similarity: 96.43% Conservative: 3
Best Local Similarity: 95.36% Mismatches: 6
Query Match: 23.51% Indels: 5
Gaps: 1
DB: 1

US-10-764-390-3 (1-1072) x AUI30115 (1-842)
Qy 347 LeuProAspAsnGluValGluLeuLysAlaPheValAlaProAlaProValGluThr 366
Db 3 TTATCCCGACAAATGAAGTTGAACCTGAAGCGCTTTGTCGCCAGCGCCACCTGTAGAACA 62
Qy 367 ThrTyHisThrGluTrpAsnLeuIleSerHisProThrAspTyTrpGlnGlyLeuLys 386
Db 63 ACCTACAACTATGATGAATTTAATAGCCACCCACAGACTACCAAGGTGAATAAATA 122
Qy 387 GlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyValPheLys 406
Db 123 CAAGGACACAAAGCAACTCTTAACCTCTCAATGTCGGTCCGACCTTATGTCTTCAA 182
Qy 407 ValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysPro 426
Db 183 GTCACTGTTCTAGTGAACCGCTTTGGAGAAGGATTTGTCAATGTCTACCTGTAAAGCCT 242
Qy 427 AlaArgValAsnLeuProProValAlaValValSerProGlnLeuGluLeuThr 446
Db 243 GCCAAGAGAGTCAACCTGCCACCTGTAGCAGTTGTTCTCCCACTGCAAGAGCTCACT 302
Qy 447 LeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleVal 466
Db 303 TTGCTTTGAGCTGAGCCCTCATTTGATGGCAGCCAAAGTACAGATGATCTGAAATAGTG 362
Qy 467 SerTyHisThrGluGluIleAsnGlyProPheIleGluGlyLysThrSerValAspSer 486
Db 363 AGTTATCATTTGGGAAGAAATAAACCGGCCCTTCATAGAAGAGAAGACTTCAGTTGACTCT 422
Qy 487 ProValLeuArgLeuSerAsnLeuAspProGlyAsnTyTrpSerPheArgLeuThrValThr 506
Db 423 CCCGCTCTTACGCTTGTCTAACCTTGTAGTGGTAACTATATAGTTTCAGGTTGACTGTACA 482
Qy 507 AspSerAspGlyAlaThrAsnSerThrAlaAlaLeuIleValAsnAsnAlaValAsp 526
Db 483 GACTCGGACGGAGCAGCACTACTCTCAACTGACGCCCTTAATAGTGAACAACTGCTGTGAC 542
Qy 527 TyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIle 546
Db 543 TACCCACAGTTGCTTAATGACGAGCAAAATCACACCATAACTTTGCCCCCAAACTCCATC 602
Qy 547 ThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuTyTrpSerLeu 566
Db 603 ACTTTGAATGAAACCAAGCAGCAGTACCAAGATTCACAGATTTGCTCTATGAGTGGTCCCTG 662
Qy 567 GlyProGlySerGluGlyLysHisValValMetGlnGlyValGlnThrProTyTrpLeuHis 586
Db 663 GGTCTGGAGTGANGGCAACATGTGTGTATGAGGGAGT-CAGACGCCATACCTTCAT 721
```

QY 587 LeuSerAlaMetGlnGluGlyAspTyrThrPheGln-LeuLysValThrAsp-SerSera 606  
 |||||  
 Db 722 TTATCTGCAATGCAGGAGGAGTATCATCTTACGTTGAGGTGACAGAACTCTTAA 781  
 |||||  
 QY 606 rgGlnGlnSerThrAla-\*\*\*ValThr-ValLleValGlnProGluuAsnAsn 622  
 |||||  
 Db 782 GGCACACAGTCTACTGCTGGGGGTGACTGNGAATGGTCCAGGCTGAAACCAT 833  
 |||||

RESULT 12  
 CB722905 710 bp mRNA linear EST 09-JUL-2003  
 LOCUS UI-M-GH0-cef-i-19-0-UI.r1 NIH\_BMAP\_GH0 Mus musculus cDNA clone  
 DEFINITION IMAGE:6838844 5', mRNA sequence.

ACCESSION CB722905  
 VERSION CB722905.1 GI:29780047

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 710)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

<http://genome.uiowa.edu/distribution/mousefl.html>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

FEATURES

source

1..710  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6838844"  
 /tissue\_type="Whole brain"  
 /dev\_stage="1, 5, and 15 days newborn"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_GH0"

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 to the protocol of Benton and Davis (1978). The library  
 was screened by hybridization to a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is CCACTGAT. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

ORIGIN

Alignment Scores:

Pred. No.:	3,07e-84	Length:	710
Score:	1051.00	Matches:	203
Percent Similarity:	92.31%	Conservative:	13
Best Local Similarity:	86.75%	Mismatches:	18
Query Match:	18.84%	Indels:	0
DB:	6	Gaps:	0

US-10-764-390-3 (1-1072) x CB722905 (1-710)

QY 575 ValValMetGlnGlyValGlnThrProTyrThrLeuHisLeuSerAlaMetGlnGluGlyAsp 594  
 |||||  
 Db 4 ATGTCATGTCAGGAGGCGCCAGACCCCATACCTTCTACTTGTCTGAGCTGCAGGAGGAGAG 63  
 |||||  
 QY 595 TyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla-\*\*\*ValThr 614  
 |||||  
 Db 64 TACACATTTACGTCGATGGTGACGGATTCCTCGGAGACCAATCCACCGCTTTGGTCGCT 123  
 |||||  
 QY 615 ValLleValGlnProGluuAsnAsnAsnProProValAlaValAlaValAlaGlyProAspLysGlu 634  
 |||||  
 Db 124 GTGACTGTCACAGGCTGAGAAACAACAGGCTCCGGTGGCGGTCGAGGAGCCGCCACAGGAG 183  
 |||||  
 QY 635 LeuLlePheProValGluSerAlaThrLeuAspGlySerSerSerSerAspAspHisGly 654  
 |||||  
 Db 184 CTGGTTTCCCGGTGCAGAGTGCTACACTGATGGAGCCGAGCAGTGACGATCACGGC 243  
 |||||  
 QY 655 IleValPheTyrHisTyrGluHisValArgGlyProSerAlaValGluMetGluAsnIle 674  
 |||||  
 Db 244 ATTGTCTGCTATCACTGGGAGCACATCAGAGGCCCCAGTCAGTCAGATGGAGAATGTG 303  
 |||||  
 QY 675 AspLysAlaLleAlaThrValThrGlyLeuGlnValGlyThrTyrThrHisPheArgLeuThr 694  
 |||||  
 Db 304 GACAAAGCCATAGCCACTGTACCAGCCCTTCAGGTGGGTATCTACCACTTCGGGTAAACC 363  
 |||||  
 QY 695 VallyAspGlnGlnGlyLeuSerSerThrThrLeuThrValAlaValLysLysGlu 714  
 |||||  
 Db 364 GTGAGACACAGCAGGCTCTGAGCAGCAGCTCCACCCTCACTGTGGCTGTGAGAGAGGA 423  
 |||||  
 QY 715 AsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAsn 734  
 |||||  
 Db 424 AATAACAGCCCTCCAGAGCCAGGCTGTCAGGAGACATGTTCTTATATCTTCCCAATAAT 483  
 |||||  
 QY 735 SerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuTyr 754  
 |||||  
 Db 484 TCCATTACTTTGGATGGTTCAGAGGTCTACTGATGACCGAGGAATGTGTCTTATCTGTGG 543  
 |||||  
 QY 755 IleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerVal 774  
 |||||  
 Db 544 ATCCGGGATGCCAGAGTCCAGCTGCCGAGATGTCAATGGAGGCTCTGACCACAGGGCA 603  
 |||||  
 QY 775 AlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrAsp 794  
 |||||  
 Db 604 GCTCTGCAGCTCACCAATCTGTCGAGAGGCGCTACACTTTTCACTTGTAGTCACTGAC 663  
 |||||  
 QY 795 SerGlnGlyAlaSerAspThrAspThrAlaThrValGluVal 808  
 |||||  
 Db 664 AGTCAGGGGGCTCCGACTCAGACGCGCCCACTGTGGAGGTG 705  
 |||||

RESULT 13  
 LOCUS AL044313  
 DEFINITION DKFZp3400729\_r1 434 (synonym: htes3) Homo sapiens cDNA clone  
 ACCESSION AL044313  
 VERSION AL044313.1 GI:5432536

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 597)

AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.

TITLE EST (Koehrer, et al.)

JOURNAL Unpublished (1999)

COMMENT Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email [s.wiemann@dkfz-heidelberg.de](mailto:s.wiemann@dkfz-heidelberg.de);

sequenced by BMFZ (Biomedical Research Center at the Charite,

Berlin/Germany) within the cDNA sequencing consortium of the German

## Genome Project.

No. s1 sequence available.  
This clone (DKFP43400729) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

Location/Qualifiers  
1..597  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFP43400729"  
/tissue="testis"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="434 (synonym: htes3)"  
/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

## ORIGIN

Alignment Scores:  
Pred. No.: 5,18e-82 Length: 597  
Score: 1025.00 Matches: 197  
Percent Similarity: 98.99% Conservative: 0  
Best Local Similarity: 98.99% Mismatches: 1  
Query Match: 18.37% Indels: 1  
DB: 1 Gaps: 0

US-10-764-390-3 (1-1072) x AL044313 (1-597)

Qy 467 SerTyrHisTrpGluGluLeuAsnGlyProPheLeuGluLysThrSerValAspSer 486  
Db 3 AGTTATCATTTGGGAAGAAATTAACGGCCCTTCATAGAAGAGAGACTTCAGTTGACTCT 62  
Qy 487 ProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThr 506  
Db 63 CCCGCTTACGCTTGCTCTAACCTTGCATCCTGTAACATAGTTTCAGGTTGACTGTTACA 122  
Qy 507 AspSerPheGlyAlaThrAsnSerThrThrAlaLeuLeuValAsnAsnAlaValAsp 526  
Db 123 GACTCGGAGGAGCCCACTAATCTACAACTGAGCCCTTAATAGTGAACAATGCTGTGGAC 182  
Qy 527 TyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIle 546  
Db 183 TACCACCAAGTTGCTAATGAGGAGCAATACACCAATACTTTGCCCCAAATCCATC 242  
Qy 547 ThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeuTyrGluTrpSerLeu 566  
Db 243 ACTTGAATGAAACACAGCAGTGACGATCACAGATTGCTCTATAGTGTGCTCCTG 302  
Qy 567 GlyProGlySerGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHis 586  
Db 303 GGTCTGGAGTGAGGGCAACATGTGCTCATGTCAGGGAGTACAGACCCATACCTTCAT 362  
Qy 587 LeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArg 606  
Db 363 TTATCTGAATGTCAGGAGGAGATTAATATTTTCAGCTGAAGTGACAGATTTCTCAAGG 422  
Qy 607 GlnGlnSerThrAla\*\*ValThrValIleValGlnProGluAsnAsnArgProProVal 626  
Db 423 CAACAGTCTACTGCTGTGCTGACTGTGATTGTCCAGCCTGAAAAACAATAGACCTCCAGTG 482  
Qy 627 AlaValAlaGlyProAspLysGluLeuPheProValGluSerAlaThrLeuAspGly 646  
Db 483 GCTGTGGCGGCGCCCTGATAAAGAGCTGATCTTCCAGTGGAAAGTGTCTACCCCTGGATGGG 542  
Qy 647 SerSerSerAspAspHisGlyIleValPheTyrHisTrp-GluHisValArg 664  
Db 543 AGCAGCAGCGATGACACGGCAATTGTCTTACCACTGCGGAGACGTCAGA 597

## RESULT 14

BP229849 598 bp mRNA linear EST 15-SEP-2004  
LOCUS BP229849 Sugano cDNA library, fetal brain Homo sapiens cDNA clone  
DEFINITION FBR04757, mRNA sequence.

## ACCESSION

BP229849

## VERSION

BP229849.1 GI:52102759

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 598)  
Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,  
Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
Genome Res. 14 (9), 1711-1718 (2004)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.

## REFERENCES

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

1..598  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="FBR04757"  
/tissue\_type="brain"  
/dev\_stage="fetal"  
/clone\_lib="Sugano cDNA library, fetal brain"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.17e-79 Length: 598  
Score: 999.00 Matches: 183  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 98.92% Mismatches: 0  
Query Match: 17.90% Indels: 0  
DB: 5 Gaps: 0

US-10-764-390-3 (1-1072) x BP229849 (1-598)

Qy 888 AsnLeuHisMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuArg 907  
Db 11 AATCTGCACATGCGCTCTCAAGAGAGAGCCGACTTCTTGTCTTTCAAGGCTCTGAGG 70  
Qy 908 ValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuThr 927  
Db 71 GTTGATACAGCAGGTGCTTCTGAAGTGTCTGCGCATGTGTCACTGCGACCCCTCACA 130  
Qy 928 LysArgCysIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAsp 947  
Db 131 AAGCGCTGATTTGCTCTCACTTATGGATGAGAGAACCTTATACAGCGTTATATCTGGAT 190  
Qy 948 GlyGluSerAsnCysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIle 967  
Db 191 GGAGAGAGCAACTGTGAGTGGAGTATATCTATGTGACAGTGTGGCTTTTACTCTTATT 250  
Qy 968 ValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLysArgThr 987  
Db 251 GTGCTAACAGAGGTTTCACTTGGCTTTGCACTGCTGCTGCAAAAGACAAAAAGGACT 310  
Qy 988 LysIleArgLysLysThrLysTyrThrIleLeuAspAsnMetAspGluGlnGluArgMet 1007  
Db 311 AAAATCAGGAAAAAACAAGTACACCATCTCTGATTAACATGGATGAACAGAAAGATG 370  
Qy 1008 GluLeuArgProLysTyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMet 1027  
Db 371 GAATCTAGGCCCAATATGTTATCAAGCCCGAGAGCAGACCAACTCCAGCTGTATG 430  
Qy 1028 ValSerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGlu 1047  
Db 431 GTATCCGAGTCTGAGTTTGACAGTGACCCAGGACACAATCTTTCAGCCGAGAGAGATGGAG 490  
Qy 1048 ArgGlyAsnProLysValSerMetAsnGlySerIleArgAsnGlyValSerPheSerTyr 1067





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 12, 2005, 10:29:21 ; Search time 1169 Seconds  
(without alignments)  
5428.540 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSLLLVTTAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10764390/runat\_12102005\_110203\_22319/app.query.fasta\_1.1223  
-DB=N\_Geneseq\_16Dec04 -QWMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10764390@cgn\_1\_1\_708@runat\_12102005\_110203\_22319 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_16Dec04:\*

1:	geneseqn1980s:*
2:	geneseqn1900s:*
3:	geneseqn2000s:*
4:	geneseqn2001as:*
5:	geneseqn2001bs:*
6:	geneseqn2002as:*
7:	geneseqn2002bs:*
8:	geneseqn2003as:*
9:	geneseqn2003bs:*
10:	geneseqn2003cs:*
11:	geneseqn2003ds:*
12:	geneseqn2004as:*
13:	geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5574	99.9	6791	13 ADR00593	Adr00593 Human 254
2	5574	99.9	6791	13 ADR00591	Adr00591 Human 254
3	5546.5	99.4	6797	13 ADR00860	Adr00860 Human 254
4	5546.5	99.4	6797	13 ADR00859	Adr00859 Human 254
5	5527	99.1	6991	13 ADR00595	Adr00595 Human 254

6	5527	99.1	6991	13	ADRO0858	
7	5358	96.0	3808	10	ADGL5055	
8	2435.5	43.6	3147	12	ADM36226	
9	2404	43.1	3846	10	ADI21796	
10	2404	43.1	3908	4	AAD11067	
11	2385	42.7	4001	10	ADF42453	
12	2293.5	41.1	3918	5	AAH89908	
13	2278	40.8	3628	10	ADGL5051	
14	2184.5	39.1	2796	4	AAH18563	
15	1817.5	32.6	2501	2	AAV58761	
16	1451.5	26.0	3345	4	ABL13095	
17	1443.5	25.9	2223	4	AAH15983	
18	1314	23.5	2423	5	AAS73181	
19	1280.5	22.9	5775	4	ABL13094	
20	845	15.1	772	4	AAH06733	
21	823	14.7	818	8	ACA56915	
22	812.5	14.6	1733	5	AAH90021	
23	681	12.2	580	2	AAV89845	
24	640	11.5	578	2	AAV89613	
C	25	477	8.5	284	13	ADRO0590
C	26	437	7.8	982	4	AAI89864
C	27	403	7.2	126990	12	ADP13332
C	28	397	7.1	1420	3	AAC69580
C	29	332	5.9	313	6	ABN22934
C	30	241.5	4.3	3612	8	ACA25573
C	31	235.5	4.2	20320	5	ABA21464
C	32	225	4.0	103052	13	ADQ89963
C	33	224	4.0	81940	4	AAS05390
C	34	224	4.0	81940	6	ABK64829
C	35	224	4.0	81940	12	ADQ17315
C	36	222.5	4.0	6614	10	ADC30653
C	37	216.5	3.9	110000	12	ADNA46845_16
C	38	216.5	3.9	110000	12	ADNA47591_04
C	39	216.5	3.9	110000	12	ADNA46123_16
C	40	216.5	3.9	110000	12	ADNA47209_04
C	41	216.5	3.9	110000	12	ADNA46464_16
C	42	216.5	3.9	110000	12	ADNA47960_04
C	43	214.5	3.8	3687	8	ABZ77634
C	44	214.5	3.8	3687	12	ADH36637
C	45	214	3.8	23545	13	ADQ89759

ALIGNMENTS

RESULT 1	
ADRO0593	
ID	ADRO0593 standard; CDNA; 6791 BP.
XX	
AC	ADRO0593;
XX	
DT	04-NOV-2004 (first entry)
XX	
DE	Human 254P1D6B v.2 encoding CDNA SEQ ID NO:4.
XX	
KW	254P1D6B; small interfering RNA; siRNA; immun
KW	254P1D6B-related protein; cytosolic; gene th
KW	254P1D6B v.2; gene; ss; chromosome 6.
XX	
OS	Homo sapiens.
XX	
FH	Location/Qualifiers
FT	512..3730
FT	/*tag= a
FT	/product= "254P1D6B v.2"
XX	
FN	WO2004067716-A2.
XX	
PD	12-AUG-2004.
XX	
PF	23-JAN-2004; 2004WO-US001965.
XX	
PR	24-JAN-2003; 2003US-0442526P.
XX	

#### ALIGNMENTS

RESULT 1	ADR00593	standard; cDNA; 6791 BP.
ID	ADR00593	
XX	ADR00593;	
AC	ADR00593;	
XX		
DT	04-NOV-2004	(first entry)
XX		
DE	Human 254P1D6B v.2	encoding cDNA SEQ ID NO:4.
XX		
KW	254P1D6B;	small interfering RNA; siRNA; immune response;
KW	254P1D6B-related protein;	cytostatic; gene therapy; cancer; human;
KW	254P1D6B v.2; gene; ss;	chromosome 6.
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	512..3730
FT		/*tag= a
FT		/product= "254P1D6B v.2"
XX		
FN	WO2004067716-A2.	
XX		
PD	12-AUG-2004.	
XX		
PF	23-JAN-2004;	2004WO-US001965.
XX		
PR	24-JAN-2003;	2003US-0442526P.
XX		

PA (AGEN-) AGENSYS INC.  
 XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
 PI Perez-Villar JJ, Paris M;  
 PI  
 DR WPI; 2004-580991/56.  
 DR P-PSDB; ADR00594.  
 XX  
 PT New 254PID6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.  
 XX  
 PS Claim 10; SEQ ID NO 4; 345pp; English.  
 PS  
 CC The present invention describes a 254PID6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of PI; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of PI; (5)  
 CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
 CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of PI; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of PI; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of PI. 254PID6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence encodes the human  
 CC 254PID6B v.2, which is used in the exemplification of the present  
 CC invention. The human 254PID6B gene is located on chromosome 6p22.  
 XX  
 SQ Sequence 6791 BP; 1812 A; 1542 C; 1684 G; 1753 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0 Length: 6791  
 Score: 5574.00 Matches: 1069  
 Percent Similarity: 99.72% Conservative: 0  
 Best Local Similarity: 99.72% Mismatches: 3  
 Query Match: 99.89% Indels: 0  
 DB: 13 Gaps: 0

US-10-764-390-3 (1-1072) x ADR00593 (1-6791)

QY 81 SerCysProHisLysGluAsnGluProLysLysMetGlyProIleArgSerTyrLeu 100  
 DB AGCTGCCCCACAAAGAGAACTGTGAGCCCAAGAGATGGGCCCATCAGGTCTTATCTC 811  
 QY 101 ThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMet 120  
 DB ACTTTTGTGCTCCGGCCCTTTCAGAGGCTTCACAGCTGTGGACTATGGGGACATGATG 871  
 QY 121 LeuAsnArgGlySerProSerGlyTyrGlyAspSerProGluAspIleArgLysAsp 140  
 DB CTGAACAGGGCTCCCTCCGGGATCTGGGGGACTCCTCCTGAGATATCAGAAAGAC 931  
 QY 141 Leu\*\*\*PheLeuGlyLysAspTyrGlyLeuGluMetSerGluTyr\*\*\*AspAspTyr 160  
 DB TTGCCCTTTCTAGCAAGATTGGGGCTAGAGAGATGTCTGAGTACGCAGATGACTAC 991  
 QY 161 ArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAla 180  
 DB CGGGAGCTGGAGAGGACTCTTTCACCCAGTGGCAAGCAGAGCCCGAGGGAGTGCC 1051  
 QY 181 GluTyrThrAspTyrGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGly 200  
 DB GAGTACAGGACTGGGGCTTACTGCCGGGCGAGCGAGGGGGCTTCAACTCTCTCTGTTGA 1111  
 QY 201 AspSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGlu 220  
 DB GACAGTCTCGGTGCCAGCGAGCGAGCGAGCCCTGAGCTCCATTACCTGAATGAG 1171  
 QY 221 SerAlaSerThrProAlaProLysLeuProGluArgSerValLeuLeuProLeuProThr 240  
 DB TCGGCTTCAACCCCTGCCCAAACTCCTCGAGAGAGTGTGTGCTTCCCTTGCCTGCGACT 1231  
 QY 241 ThrProSerSerGlyGluValLeuGluLysGluLysAlaSerGlnLeuGluGlnSer 260  
 DB ACTCATCTTTCAGAGAGGTGTGGAGAAAGAAAGGCTTCTCAGCTCCAGACAAATCC 1291  
 QY 261 SerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProAlaSerLeu 280  
 DB AGCAACAGCTCTGGAAAGAGTCTTAATGCTTCCCATAGTCTTCTCCCGCAAGCCTG 1351  
 QY 281 GluLeuSerSerValThrValGluLysSerProValLeuLeuThrValThrProGlySerThr 300  
 DB GAGCTCAGCTCAGTCACCGTGAGAAAAAGCCAGTGTCTCAGAGTCAACCCCGGGAGTACA 1411  
 QY 301 GluHisSerIleProThrProThrSerAlaAlaProSerGluSerGluSerProSerGlu 320  
 DB GAGCAGCATCTCCCAACACCTCCCACTAGCGAGGCCCCCTCTGAGTCCACCCCATCTGAG 1471  
 QY 321 LeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAlaGly 340  
 DB CTACCCATATCTCTTACCTACCTGCTCCAGAGGACAGTGAAGAACTTACGGTATCGGCTGA 1531  
 QY 341 AspAsnLeuIleIleThrLeuProAspAsnGluValGluLeuLysAlaPheValAlaPro 360  
 DB GATTAACCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1591  
 QY 361 AlaProProValGluThrThrTyrAsnTyrGluTyrAsnLeuIleSerHisProThrAsp 380  
 DB GCGCACCTGTAGAAACCAACCTTACAACTATGAATGAATGAATGAATGAATGAATGAATGA 1651  
 QY 381 TyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerVal 400  
 DB TACCAAGGTGAATAAATAAACAAGGACACAAAGCAAACTCTTAACCTCTCTCAATGTCCGTC 1711  
 QY 401 GlyLeuTyrValPheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheVal 420  
 DB GGACTTTATGTCTTCAAAAGTCACTGTTTCTAGTGAAGGCGCCCTTTCGAGAGAGGATGTC 1771  
 QY 421 AsnValThrValLysProAlaArgValLeuLeuProValAlaValValSerPro 440  
 DB AATGTCACCTGTAAAGCCCTGCCAGAGAGTCAACCTTCCACCTTGTAGCAGTGTGTTCTCCC 1831  
 QY 441 GlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuAspGlySerGlnSerThr 460

Db 1832 CAATGCAAGAGCTCATTGCGCTTTGACGTCAGCCCTCATTTGATGGGAGCCAAAGTACA 1891  
Qy 461 AspAspThrGluIleValSerTyrHisTrpGluIleAenGlyProPheIleGluGlu 480  
Db 1892 GATGATACTGAATAGTAGTATCAATGGGAAGATAAAGCGGCCCTTCATAGAGAG 1951  
Qy 481 LysThrSerValAspSerProValLeuArgLeuSerAenLeuAspProGlyAsnTyrSer 500  
Db 1952 AAGACTTCAGTTGACTCTCCGCTTACGCTTGCTAACCTTGATCTCGTAACTATAGT 2011  
Qy 501 PheArgLeuThrValThrAspSerAspGlyAlaThrAenSerThrAlaAlaLeuLe 520  
Db 2012 TTCAGGTTGACTGTTACAGACTCCGACGGACCCACTAACTCTCAACTGCGGCCCTAATA 2071  
Qy 521 ValAenAenAlaValAspTyrProProValAlaAenAlaGlyProAenHisThrIleThr 540  
Db 2072 GTGAACAATGCTGTGGACTACCCACCACTTGTCTAATGCAGGACCAATCAACCATTA 2131  
Qy 541 LeuProGlnAenSerIleThrLeuAenGlyAenGlnSerSerAspAspHisGlnIleVal 560  
Db 2132 TTGCCCAAACTCCATCACTTTGAATGGAAACAGAGCAGTGCAGATCAACAGATTGTC 2191  
Qy 561 LeuTyrGluTrpSerLeuGlyProGlySerGluGlyLysHisValMetGlnGlyVal 580  
Db 2192 CTCTATGAGTGGTCCCTGGGTCCTGGGAGTGAGGCGCAACATGTGGTCATGCGGGAGTA 2251  
Qy 581 GlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLys 600  
Db 2252 CAGAGCCCAATCTTCATTTATCTGCAATGCAGGAGGAGATATACATTTACGTCGAG 2311  
Qy 601 ValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnProGlu 620  
Db 2312 GTGACAGATTCTTCAAGCGCAACAGCTACTCTGCTAGTACTGTGATTGTCCAGCCTGAA 2371  
Qy 621 AsnAenArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGlu 640  
Db 2372 AACAAATAGACTCCAGTGGCTGTGGCGGCCCTCATAAAGAGCTGATCTTCCCAAGTGAA 2431  
Qy 641 SerAlaThrLeuAspGlySerSerSerSerAspAspHisGlyIleValPheTyrHisTrp 660  
Db 2432 AGTCTACCTCGATGGGAGCGAGCGAGCGATGACCAAGCATTTCTTACCACTGG 2491  
Qy 661 GluHisValArgGlyProSerAlaValGluMetGluAenIleAspLysAlaIleAlaThr 680  
Db 2492 GAGCAGCTCAGAGGCCCCAGTGCAGTGAGATGAAATAATTGACAAAGCAATAGCCACT 2551  
Qy 681 ValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGly 700  
Db 2552 GTGACTGGTCTCCAGGTGGGAGACCTACCACTTCCGTTTGACAGTGAAAGACCAAGCGGA 2611  
Qy 701 LeuSerSerThrSerThrLeuThrValAlaValLysLysGluAenAenSerProArg 720  
Db 2612 CTGAGCAGCAGCTCCACCTCAGCTGTGGCTGTGAAGAGGAAATATAGTCTCCAGAG 2671  
Qy 721 AlaArgAlaGlyArgHisValLeuValLeuProAenAenSerIleThrLeuAspGly 740  
Db 2672 GCCCGGGCTGGTGAGCAGATGTTCTTGTGCTTCCCAATATTCATTACTTTGGATGGT 2731  
Qy 741 SerArgSerThrAspAspGlnArgIleValSerTyrLeuTrpIleArgAspGlyGlnSer 760  
Db 2732 TCAAGGTCTACTGATGACCAAGAAATTTGTCTCTATCTGTGGATCCGGGATGGCCAGAGT 2791  
Qy 761 ProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAen 780  
Db 2792 CCAGCAGCTGGAGATGTATCATGGCTCTGACACAGTGTGGCTCTGCAGCTTACGAAT 2851  
Qy 781 LeuValGluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAsp 800  
Db 2852 CTGGTGGAGGGGTGTACACTTTCACCTTGGAGTCCCGACAGTCAGGGGGCCCTCGGAC 2911  
Qy 801 ThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValGluLeu 820

Db 2912 ACAGACACTCCCACTGTGGAGAGTGCAGCCAGACCCCTAGGAGAGTGCCTGGTGGAGCTG 2971  
Qy 821 ThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAspThrLeuValArgGln 840  
Db 2972 ACCCTGCAGGTGGTGTGGGAGCTGACAGAGCAGCGGAGGACACCCCTTGTGAGGCAG 3031  
Qy 841 LeuAlaValLeuLeuAenValLeuAspSerAspIleLysValGlnLysIleArgAlaHis 860  
Db 3032 CTGGCTGTGTCTGTAACGCTGGACTCGGACATTAAAGGTCGAGAAGATTCGGGGCCAC 3091  
Qy 861 SerAspLeuSerThrValIleValPheTyrValGlnSerArgProProPheLysValLeu 880  
Db 3092 TCGGATCTCAGCACCGGTGATTTGTTTATGTACAGAGCAGCGCCCTTCAAGGTTCTC 3151  
Qy 881 LysAlaAlaGluValAlaArgAenLeuHisMetArgLeuSerLysGluLysAlaAspPhe 900  
Db 3152 AAAGCTGCTGAAGTGGCCGAAATCTGCACATGCGGCTCTCAAAGGAGAAAGGCTGACTTC 3211  
Qy 901 LeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHis 920  
Db 3212 TTGCTTTTCAAGGTCTTTGAGGGTTGTATACAGCAGGTTGCCCTTCTGAAAGTGTCTGGCCAT 3271  
Qy 921 GlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAenLeu 940  
Db 3272 GGTCACTGCCACCCCTCACAAAGCGCTGCATTTGCTCTCACTTATCGATGGAGAACCTT 3331  
Qy 941 IleGlnArgTyrIleTrpAspGlyGluSerAenCysGluTrpSerIlePheTyrValThr 960  
Db 3332 ATACAGCGTTATATCTGGGATGGAGAGAGCACTGTGAGTGGAGTATATTCTATGTGACA 3391  
Qy 961 ValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCys 980  
Db 3392 GTGTGGCTTTTACTCTTATTGTGCTAACAGGAGGTTTCACTTGGCTTTGCATCTGCTGC 3451  
Qy 981 CysLysArgGlnLysArgThrLysIleArgLysLysThrLysTrpThrIleLeuAenAsn 1000  
Db 3452 TGCANAAAGACAAAAGAGACTAAATCAGNAAAAAACAAGTACACCATCTCGATTAAC 3511  
Qy 1001 MetAspGlnGluArgMetGluLeuArgProLysTyrGlyIleLysHisArgSerThr 1020  
Db 3512 ATGGATGAACAGGAAAGATGGAACTGAGGCCCAATATGGTATCAAGCAGCAGGACACA 3571  
Qy 1021 GluHisAenSerSerLeuMetValSerGluSerGluPheAspSerAspGlnAspThrIle 1040  
Db 3572 GAGCACAACCTCCAGCCTGATGGTATCCGAGTCTGAGTTTGACAGTGACCGAGGACACAATC 3631  
Qy 1041 PheSerArgGluLysMetGluArgGlyAenProLysValSerMetAenGlySerIleArg 1060  
Db 3632 TTCAGCCGAGAAAGATGGAGAGGGAATCCAAAGGTTTCCATGATGGTTCCATCAGA 3691  
Qy 1061 AsnGlyAlaSerPheSerTyrCysSerLysAspArg 1072  
Db 3692 AATGGAGCTTCCCTTCACTTATTTGCTCAAGAGGACAGA 3727

RESULT 2  
ADRD00591  
ID ADR00591 standard; cDNA; 6791 BP.  
XX AC ADR00591;  
XX DT 04-NOV-2004 (first entry)  
XX DE Human 254P1D6B v.1 clone LCP-3 encoding cDNA SEQ ID NO:2.  
XX KW 254P1D6B; small interfering RNA; siRNA; immune response;  
XX KW 254P1D6B-related protein; cytosolic; gene therapy; cancer; human;  
XX OS Homo sapiens.  
XX PH Location/Qualifiers  
XX FT 512..3730  
XX FT /\*tag= a



```
Qy 381 TyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAenLeuSerGlnLeuSerVal 400
Db 1652 TACCAAGGTGAATATAAAACAAGGACACAAAGCAAACTCTTAACCTCTCTCAATTTGCGGTC 1711
Qy 401 GlyLeuTyrValPheLysValThrValSerSerGluAenAlaPheGlyGluGlyPheVal 420
Db 1712 GGACTTATGCTCTCAAAAGTCACTGTTTCTAGTGGAAAGCGCTTTGGGAAGGATTTGTC 1771
Qy 421 AsnValThrValLysProAlaArgValAsnLeuProProValAlaValValSerPro 440
Db 1772 AATGTCACTGTAAGCCTGCCAGAAGAGTCAACCTGCCCTCATTTGATGGCAGCAAGTACA 1831
Qy 441 GlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThr 460
Db 1832 CAATGCAAGAGCTCACTTTGCCCTTGAAGTCAAGCTGAGCTGAGCAGTGTGTTCTCCC 1891
Qy 461 AspAspThrGluIleValSerTyrHisTrpGluGluLeuAenGlyProPheIleGlu 480
Db 1892 GATGATACTGAAATAGTGAATATCAATGGGAAGAAATAAACGGGCCCTTCATAGAGAG 1951
Qy 481 LysThrSerValAspSerProValLeuArgLeuSerAenLeuAspProGlyAsnTyrSer 500
Db 1952 AAGACTTCAGTTGACTCTCCCGTCTTACGCTTGCTTAACCTTTGATCTGTTAACTATAGT 2011
Qy 501 PheArgLeuThrValThrAspSerAspGlyAlaThrAenSerThrThrAlaAlaLeuIle 520
Db 2012 TTCAAGTTGACTGTTTACAGACTCGGAGGAGCCCACTAACTCTACAACCTGCAGCCCTAATA 2071
Qy 521 ValAenAenAlaValAspTyrProProValAlaAenAlaGlyProAenHisThrIleThr 540
Db 2072 GTGAACAATGCTGTGGACTACCCACGATTCTTAATGAGGAGCAAAATACACCATTAAT 2131
Qy 541 LeuProGlnAenSerIleThrLeuAenGlyAsnGlnSerSerAspHisGlnIleVal 560
Db 2132 TTGCCCCAAACTCCATCACTTTGAATGGAACACAGAGCAGTGAAGTACACAGATTGTC 2191
Qy 561 LeuTyrGluTrpSerLeuGlyProGlySerGlnGlyLysHisValMetGlnGlyVal 580
Db 2192 CTCTATAGTGGTCCCTGGGTCCTGGGAGTGAAGGCAAAACATGTGTGCATGACGGAGTA 2251
Qy 581 GlnThrProTyrLeuHisLeuSerAlaMetGlnGlnGlyAspTyrThrPheGlnLeuLys 600
Db 2252 CAGACGCATACCTTCATTATCTGCAATGCAAGAGGAGATATATACATTTTCACTGAAG 2311
Qy 601 ValThrAspSerSerArgGlnGlnSerThrAla**ValThrValIleValGlnProGlu 620
Db 2312 GTGACAGATTCTTCAAGGCAACAGCTACTCTGTGTGACTGTGTATGTGCCACCTGAA 2371
Qy 621 AsnAenArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGlu 640
Db 2372 AACAAATAGACCTCCAGTGGCTGTGGCGGCCCTGATAAAGAGCTGATCTTCCAGTGGAA 2431
Qy 641 SerAlaThrLeuAspGlySerSerSerAspAspHisGlyIleValPheTyrHisTrp 660
Db 2432 AGTCTACCTCCGTGGATGGAGCAGCAGCAGCATGACCAAGCATTTGTCTTACACACTGG 2491
Qy 661 GluHisValArgGlyProSerAlaValGluMetGluAenIleAspLysAlaIleAlaThr 680
Db 2492 GAGCAGTCAAGGCCCCAGTGCAGTGGAGATGGAAATATATGACAAAGCAATAGCCACT 2551
Qy 681 ValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGly 700
Db 2552 GTGACTGGTCTCCAGGTGGGACCTTACCCTTCGTTTGCAGTGAAAGACCAAGGGA 2611
Qy 701 LeuSerSerThrSerThrLeuThrValAlaValLysLysGluAenAenSerProProArg 720
Db 2612 CTGAGCAGCAGCTCCACCCCTCACTGTGGCTGTGAAGAGGAAATAATAGTCTCTCCAGA 2671
Qy 721 AlaArgAlaGlyGlyArgHisValLeuValLeuProAenAenSerIleThrLeuAspGly 740
Db 2672 GCCCGGGCTGGTGACAGACATGTTCTGTGCTTCCCAATAATTCATTTCAATTTGATGGT 2731
```

```
Qy 741 SerArgSerThrAspAspGlnArgIleValSerTyrIleuTrpIleArgAspGlyGlnSer 760
Db 2732 TCAAGGTCCTACTGATGACCAAGAAATTTGTCTTATCTGTGATCCGGGATGGCCAGAGT 2791
Qy 761 ProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAen 780
Db 2792 CCACGAGCTGGAGATGTCATCGATGGCTCTGACCCACAGTGTGGCTCTGCAGCTTACGAAT 2851
Qy 781 LeuValGluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAsp 800
Db 2852 CTGTGGAGGGGGTGTACACTTCCACTTGGAGTCCACCGACAGTCAGGGGGCTCTGGAC 2911
Qy 801 ThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValGluLeu 820
Db 2912 ACAGACACTGCGACTGTGGAGTGCAGCCACTAGGAAGTGGGCTCTGGTGGAGCTG 2971
Qy 821 ThrLeuGlnValGlyValGlnLeuThrGluGlnArgLysAspThrLeuValArgGln 840
Db 2972 ACCCTGCAGGTTGGTGTGGGACCTGCAGAGCAGCGGAAGGACACCTTGTGAGGAG 3031
Qy 841 LeuAlaValLeuLeuAenValLeuAspSerAspIleLysValGlnLysIleArgAlaHis 860
Db 3032 CTGGCTGTGCTGCTGAACGCTGCGACTCGGACATTAAGGTCAGAGATTCGGGCCAC 3091
Qy 861 SerAspLeuSerThrValIleValPheTyrValGlnSerArgProPheLysValLeu 880
Db 3092 TCGGATCTCAGCACCGTGATTTGTTTATGTACAGAGCAGCGGCCCTTTCAAGGTTCTC 3151
Qy 881 LysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAspPhe 900
Db 3152 AAAGCTGCTGAGTGGCCCCGAAATCTGCACATCGGCTCTCAAGGAGAAAGCTGACTTC 3211
Qy 901 LeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHis 920
Db 3212 TTGCTTTTCAAGGTCCTTGAAGGTTGATACAGAGGTTGCCCTCTCTGAAGTGTCTGGCCAT 3271
Qy 921 GlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAenLeu 940
Db 3272 GGTCACTCGCAGCCCCCTCACAAAGCGCTGCATTTGCTCTCTCATTTAGATGGAGAACCTT 3331
Qy 941 IleGlnArgTyrIleTrpAspGlyGluSerAsnCysGluTrpSerIlePheTyrValThr 960
Db 3332 ATACAGCGTTATATCTGGGATGGAGAGAGCAACTGTGAGTGGAGTATATTTCTATGTGACA 3391
Qy 961 ValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCys 980
Db 3392 GTGTTGGCTTTTACTTCTTATTGCTTAACAGAGGTTTCACTTGGCTTTGCATCTGCTGC 3451
Qy 981 CysLysArgGlnLysArgThrLysIleArgLysLysThrLysTyrThrIleLeuAspAen 1000
Db 3452 TGCAAAAGACAAAAAAGGACTAAATAATCAGGAAAAAAACAAAGTACACCATCTCGATAAC 3511
Qy 1001 MetAspGluGlnGluArgMetGluLeuArgProLysTyrGlyIleLysHisArgSerThr 1020
Db 3512 ATGATGGAACAGGAAAGAAATGGAACCTGAGGCCCAAAATATGTTATCAAGCCGAGGACA 3571
Qy 1021 GluHisAsnSerSerLeuMetValSerGluSerGluPheAspSerAspGlnAspThrIle 1040
Db 3572 GAGCACAACCTCCAGCTGATGGTATCCGAGTCTGAGTTTGACAGTGAAGCAGCAGCAATC 3631
Qy 1041 PheSerArgGluLysMetGluArgGlyAsnProLysValSerMetAsnGlySerIleArg 1060
Db 3632 TTCAGCCGAGAAAAAGATGGAGAGAGGGAATCCAAAGGTTTCCATGAATGGTTCCATCAGA 3691
Qy 1061 AsnGlyAlaSerPheSerTyrCysSerLysAspArg 1072
Db 3692 AATGAGGCTTCCCTTCAGTTATTTGCTCAAGAGACAGA 3727
RESULT 3
ADRO0860
ID ADRO0860 standard; DNA; 6797 BP.
XX
AC ADRO0860;
```

XX 04-NOV-2004 (first entry)  
 XX Human 254P1D6B v.3 nucleotide sequence SEQ ID NO:271.  
 XX 254P1D6B; small interfering RNA; siRNA; immune response;  
 KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254P1D6B v.3; chromosome 6; gene; ds.  
 XX Homo sapiens.  
 OS  
 XX WO2004067716-A2.  
 PN  
 XX 12-AUG-2004.  
 PD  
 XX 23-JAN-2004; 2004WO-US001965.  
 PF  
 XX 24-JAN-2003; 2003US-0442526P.  
 PR  
 XX (AGEN-) AGENSYS INC.  
 PA  
 XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;  
 XX WPI; 2004-580991/56.  
 XX  
 XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.  
 XX  
 XX Example 5; SEQ ID NO 271; 345bp; English.  
 PS  
 XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of PI; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of PI; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of PI; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of PI; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of PI. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254P1D6B v.3 nucleotide sequence, which is used in the  
 CC exemplification of the present invention. The human 254P1D6B gene is  
 CC located on chromosome 6p22.  
 XX  
 XX Sequence 6797 BP; 1812 A; 1542 C; 1686 G; 1757 T; 0 U; 0 Other;  
 SQ

Alignment Scores:  
 Pred. No.: 0 Length: 6797  
 Score: 5546.50 Matches: 1069  
 Percent Similarity: 99.53% Conservative: 0  
 Best Local Similarity: 99.53% Mismatches: 3  
 Query Match: 99.40% Indels: 2  
 DB: 13 Gaps: 1

US-10-764-390-3 (1-1072) x ADR00860 (1-6797)

QY 1 MetAlaProThrGlyValLeuSerSerLeuLeuValThrIleAla---G1 19  
 DB 514 ATGGCGCCCCCAGAGGTGCTCTCTTCATTGTCTGCTGGTGACAAATTCAGTTGG 573  
 QY 19 YCvAlaArgLysGlnCysSerGluGlyArgThrTy-SerAsnAlaValIleSerProAs 39  
 DB 574 TTGTCCCGTAAGCAGTCGACGAGGGGAGACATATTCATATGCAGTCATTTCACCTAA 633  
 QY 39 nLeuGluThrThrArgIleMetArgValSerHisThrPheProValValAlaSerCysThrAl 59  
 DB 634 CTTGGAACACACAGAAATCATGCGGTGCTCTCACACCTTCCTGTGTGCTAGACTGCACGCG 693  
 QY 59 aAlaCysCysAspLeuSerSerCysAspLeuAlaTrrPheGluGlyArgCysThrIle 79  
 DB 694 CGCTTGTGTGACCTGTCCAGCTGTGACCTGGCTGGTGGTTCGAGGGCGCTGTCTACTCT 753  
 QY 79 uValSerCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTy 99  
 DB 754 GGTGAGCTGCCCCCACAAGAGAACTGTGAGCCCCCAAGAGATGGGCCCCATCAGTCTTA 813  
 QY 99 rLeuThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMe 119  
 DB 814 TCTCATTCTTGTGCTCGGCTGTTCAGAGGCTGACAGCTGTCTGGACTATGGGACAT 873  
 QY 119 tMetLeuAsnArgGlySerProSerGlyIleTrrPheGlyAspSerProGluAspIleArgly 139  
 DB 874 GATGCTGAACAGGGGCTCCCTCGGGGATCTGGGGGACTCACCTGAGGATATCAGAAA 933  
 QY 139 sAspLeu\*\*\*PheLeuGlyLysAspTrrPheGlyLeuGluMetSerGluTy\*\*\*AspAs 159  
 DB 934 GGACTTGGCTTTCTAGGCAAGATTGGGGCTAGAGGAGATGCTGTAGTACTCAGATGA 993  
 QY 159 pTyArgGluLeuGlyAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySe 179  
 DB 994 CTACGGGAGCTGGAGAGGACCTCTTGCAACCCAGTGGCAAGCAGAGGAGCCAGAGGGAG 1053  
 QY 179 rAlaGluTyThrAspTrrPheGlyLeuLeuProGlySerGluGlyAlaPheAsnSerVa 199  
 DB 1054 TGCCGAGTACACGAGCTGGGGCTTACTCCGGGAGCAGGGGGCTTCAACTCTCTGT 1113  
 QY 199 lGlyAspSerProAlaValProAlaGluThrGlnAspProGluLeuHisTyLeuAs 219  
 DB 1114 TGGAGACAGTCTCTGGGTGCCAGCGAGACGACGAGGAGCCCTGAGCTCCATTACTGAA 1173  
 QY 219 nGluSerAlaSerThrProAlaProLysLeuProGluArgSerValLeuLeuProLeuPr 239  
 DB 1174 TGAGTGGGCTTCAACCCCTGCCCCCAAACTCCCTGAGAGAAAGTGTGTGTCTTCCCTTGGC 1233  
 QY 239 oThrThrProSerSerGlyGluValLeuGluLysGluLysAlaSerGlnLeuGlnGlu 259  
 DB 1234 GACTACTCCATCTTCAGGAGAGGTTCATATGCTTCCCATAGTCTTCCCTCCGCAAG 1293  
 QY 259 nSerSerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProAlaSe 279  
 DB 1294 ATCCAGCACAGCTCTGGAAAGAGGTTCTAATGCTTCCCATAGTCTTCCCTCCGCAAG 1353  
 QY 279 rLeuGluLeuSerSerValThrValGluLysSerProValLeuThrValThrProGlySe 299  
 DB 1354 CCTGAGGCTCAGCTCAGTCAGTCAGTGGAGAAAGCCAGTGTCTCACAGTCAACCCCGGGAG 1413  
 QY 299 rThrGluHisSerIleProThrProProThrSerAlaAlaProSerGluSerThrProSe 319  
 DB 1414 TACAGCACAGCATCCCAACACCTCCCACTAGCGAGCCCCCTCTGAGTCCACCCCATC 1473  
 QY 319 rGluLeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAl 339  
 DB 1474 TGAGCTACCCATATCTCTACACCTGCTCCCGAGCAGTGAAGAACTTACGGTATCGGC 1533  
 QY 339 aGlyAspLeuLeuIleThrLeuProAspLeuValGluLeuLysAlaPheValAl 359



Db 1534 TGGAGATAACCTAATTATTAACCTTACCAGCAATGAAGTTGAAGGCGCTTTGTTGC 1593  
Qy 359 aProAlaProValGluThrThrTyAsnTyxGluTrpAsnLeuLeuSerHisProTh 379  
Db 1594 GCCAGCGCCACTGTAGAAAACAACCTAACAATGATGGAATTAATAAGCCACCCAC 1653  
Qy 379 rAspTyxGlnGlyGluLeuLeuGlyHisGlyGlnThrLeuAsnLeuSerGlnLeuSe 399  
Db 1654 AGACTACCAAGGTGAATAAACAAGGACACAAGCAAACTCTTAACCTCTCTCAATTTGTC 1713  
Qy 399 rValGlyLeuTyxValPheLeuValThrValSerSerGluAsnAlaPheGlyGluGlyPh 419  
Db 1714 CGTCGGACTTATGTCTTCAAGTCACTGTTCTAGTGAACCGCTTTGGAGAGGAT 1773  
Qy 419 eValAsnValThrValLysProAlaArgValAsnLeuProProValAlaValValSe 439  
Db 1774 TGTCAATGTCACTGTTAAGCGTCGAGAAAGTCAACCTGCCACCTGTAGCAGTTGTTTC 1833  
Qy 439 rProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuLeuLeuAspGlySerGlnSe 459  
Db 1834 TCCCAACTGCAAGAGCTCACTTTGCGCTTTGACGTCAGCCCTCAATTGATGGACCCAAAG 1893  
Qy 459 rThrAspAspThrGluLeuValSerTyxHisTrpGluGluLeuAsnGlyProPheIleGl 479  
Db 1894 TACAGATGATCTGAATAGTAGTTATCATTTGGAGAGAAATAAACGGCCCTTCATAGA 1953  
Qy 479 uGluLysThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTy 499  
Db 1954 AGAGAAGACTTTCAGTTGACTCTCCGCTTTCAGCTTGTCTAACTTGATCTGGTAACTA 2013  
Qy 499 rSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaLe 519  
Db 2014 TAGTTTCAGGTTGACTGTGTACAGACTCGGACGGAGCCCACTAACTCTACACTGCGAGCCCT 2073  
Qy 519 uIleValAsnAsnAlaValAspTyxProProValAlaAsnAlaGlyProAsnHisThrIl 539  
Db 2074 AATAGTGAACTAGTGTGGACTACCCACAGTTGCTTAATCAGGACCAATACACANT 2133  
Qy 539 eThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIl 559  
Db 2134 AACTTTGCCCAAACTCCATCACTTTGATGGAACCCAGAGCAGTACAGATCACCAGAT 2193  
Qy 559 eValLeuTyxGluTrpSerLeuGlyProGlySerGluGlyLysHisValValMetGlnGl 579  
Db 2194 TGTCTCTATAGTGTGTCCTGGGTCTGGAGTGGAGGCAAAATGTTGGTCAATGAGGG 2253  
Qy 579 yValGlnThrProTyxLeuHisLeuSerAlaMetGlnGluGlyAspTyxThrPheGlnLe 599  
Db 2254 AGTACAGCGCATACCTTCACTTTATCTGCAATGCAAGAGAGATTATACATTCAGCT 2313  
Qy 599 uLysValThrAspSerSerArgGlnGlnSerThAla\*\*\*ValThrValIleValGlnPr 619  
Db 2314 GAAGTGACAGATCTTCAAGGCAACAGTCTACTGCTGTGTGATGTTGTCAGGCC 2373  
Qy 619 oGluAsnAsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProVa 639  
Db 2374 TGAANAACAATAGACCTCCAGTGGCTGTGGCGGCCCTGATAAAGAGCTGATCTTCCCACT 2433  
Qy 639 lGluSerAlaThrLeuAspGlySerSerSerAspAspHisGlyIleValPheTyxHi 659  
Db 2434 GGAAGTGCTACCTTGGATGGAGCAGCAGCAGCGATGATCACCAGCGATTTGCTTCTACCA 2493  
Qy 659 eTrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAl 679  
Db 2494 CTGGGAGCACGTGAGAGGCCCGCAGTGCAGTGGAGATGGAAATATTGACAAAGCAATAGC 2553  
Qy 679 aThrValThrGlyLeuGlnValGlyThrTyxHisPheArgLeuThrValLysAspGlnGl 699  
Db 2554 CACTGTGACTGTCTCCAGTGGGACCTTACCCTTTCGCTTTCAGTGAAGAACGACCA 2613  
Qy 699 nGlyLeuSerSerThrThrLeuThrValAlaValLysLysGluAsnAsnSerProPr 719  
Db 2614 GGGAGTGAAGCAGCAGTCCACCTCACCTGTGGCTGTGAAGAGGAAATAATAGTCTCTCC 2673

Qy 719 oArgAlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAs 739  
Db 2674 CAGAGCCCGGCTGTGGCAGACATGTTCTTGTGCTTCCCAATAATTCCATTACTTTGGA 2733  
Qy 739 pGlySerArgSerThrAspAspGlnArgIleValSerTyxLeuTrpIleArgAspGlyGl 759  
Db 2734 TGGTTCAGGTCTACTGATGACCAAGANTTGTCTCTATCTGTGGATCCGGATGGCA 2793  
Qy 759 nSerProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuTh 779  
Db 2794 GAGTCCAGCAGCTGGAGATGTCACTGATGGCTGTGACCACAGTGTGGCTCTGCAGCTTAC 2853  
Qy 779 rAsnLeuValGluGlyValTyxThrPheHisLeuArgValThrAspSerGlnGlyAlaSe 799  
Db 2854 GAATCTGTGGAGGGGTGTACACTTTTCCACTTCCGAGTCCACGACAGTCCAGGGGGCTC 2913  
Qy 799 rAspThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValGl 819  
Db 2914 GGACACAGACTCCCACTGTGGAAAGTGACCCAGACCTTAGGAAGAGTGGCTTGGTGA 2973  
Qy 819 uLeuThrLeuGlnValGlyValGlnLeuThrGluGlnArgLysAspThrLeuValAr 839  
Db 2974 GCTCACCTCTGCAGTGTGGTGTGGGAGCTGACAGAGCAGCGGAAGGACACCTTGTGAG 3033  
Qy 839 gGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLysValGlnLysIleArgAl 859  
Db 3034 GCAGCTGGCTGTGTGCTGAACGTGTGGACTCGGACATTAAGGTCCCAAGAGATTCGGGC 3093  
Qy 859 aHisSerAspLeuSerThrValIleValPheTyxValGlnSerArgProPheLysVa 879  
Db 3094 CCCTCGGATCTCAGCACCGGTGATGTGTTTATGTACAGAGCAGCGCCGCTTTCAGGT 3153  
Qy 879 lLeuLysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAs 899  
Db 3154 TCTCAAGCTGTGAGTGGGCCGGAATCTGCACATGCGGCTCTCAAGAGGAGAGGCTGA 3213  
Qy 899 pPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGl 919  
Db 3214 CTTCTTGTCTTCAAGGTCTTGGAGTGTGATACAGCAGGTTCGCTTCTTGAAGTGTCTGG 3273  
Qy 919 yHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAs 939  
Db 3274 CCATGGTCATGCGACCCCTCACAAAGCGCTGGATTTGCTCTCACTTATGATGGAGAA 3333  
Qy 939 nLeuIleGlnArgTyxIleTrpAspGlyGluSerAsnCysGluTrpSerIlePheTyxVa 959  
Db 3334 CCTTATACAGCGTTATATCTGGGATGGAGAGCAACTGTGAGTGGAGTATATCTATGT 3393  
Qy 959 lThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCy 979  
Db 3394 GACAGTGTGGCTTTTACTCTTATTTGTCTAACAGGAGTTTCACTTGGCTTTGCACTGTG 3453  
Qy 979 sCysCysLysArgGlnLysArgThrLysIleArgLysLysThrLysTyxThrIleLeuAs 999  
Db 3454 CTGCTGCAAAAGACAAAAGAGACTAAATCAGGAAAAAACAAGTACACCATCTCTGGA 3513  
Qy 999 pAsnMetAspGluGlnGluArgMetGluLeuArgProLysTyxGlyIleLysHisArgSe 1019  
Db 3514 TAAACATGGATGAACAGGAAGAAATGGAACCTGAGGCCCAATAATGATATCAAGCACCAG 3573  
Qy 1019 rThrGluHisAsnSerSerLeuMetValSerGluSerGluPheAspSerAspGlnAspTh 1039  
Db 3574 CACAGACACAACCTCCAGCGCTGATGGTATCCGAGTCTCGAGTTGAGTTCACAGTCCAGGAC 3633  
Qy 1039 rIlePheSerArgGluLysMetGluArgGlyAsnProLysValSerMetAsnGlySerIl 1059  
Db 3634 AATCTTCAGCGGAAAGAGTGGAGAGGGAATCCAAAGGTTTCCATGAATGTTCCAT 3693  
Qy 1059 eArgAsnGlyAlaSerPheSerTyxSerLysAspArg 1072  
Db 3694 CAGAAATGGAGCTTCTTCTCAGTTATTTGCTCAAGGACAGA 3733

RESULT 4	Score:	5546.50	Matches:	1069
ADRO0859	Percent Similarity:	99.53%	Conservative:	0
ID ADRO0859 standard; DNA; 6797 BP.	Best Local Similarity:	99.53%	Mismatches:	3
XX	Query Match:	99.40%	Indels:	2
AC ADRO0859;	DB:	13	Gaps:	1
XX	US-10-764-390-3 (1-1072) x ADRO0859 (1-6797)			
DT DT				
XX	Human 254PID6B v.1 nucleotide sequence SEQ ID NO:270.			
DE				
XX	254PID6B; small interfering RNA; siRNA; immune response;			
KW	254PID6B-related protein; cytotoxic; gene therapy; cancer; human;			
KW	254PID6B v.1; chromosome 6; gene; ds.			
XX	Homo sapiens.			
OS				
XX	WO2004067716-A2.			
PN				
XX	12-AUG-2004.			
XX	23-JAN-2004; 2004WO-US001965.			
PF				
XX	24-JAN-2003; 2003US-0442526P.			
PR				
XX	(AGEN-) AGENSYS INC.			
PA				
XX	Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;			
PI	Perez-Villar JJ, Faris M;			
PI				
XX	WPI; 2004-580991/56.			
XX				
PT	New 254PID6B siRNA composition comprising a double stranded siRNA that			
PT	corresponds to the nucleic acid ORF sequence which encodes the 254PID6B			
PT	protein or corresponds to a subsequence of the ORF, useful for detecting			
PT	and treating cancer.			
XX				
PS	Example 5; SEQ ID NO 270; 345pp; English.			
XX				
CC	The present invention describes a 254PID6B small interfering RNA (siRNA)			
CC	composition that comprises a double stranded siRNA that corresponds to			
CC	the nucleic acid open reading frame (ORF) sequence which encodes the			
CC	254PID6B protein, or corresponds to a subsequence of the ORF, where the			
CC	double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous			
CC	nucleotides in length. Also described: (1) a composition that comprises,			
CC	consists essentially of, or consists of a peptide of eight, nine, ten, or			
CC	eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID			
CC	NO:3, 5 or 7 ADRO0592, ADRO0594 or ADRO0596), or a peptide included in			
CC	any of the 42 lists of peptides, given in the specification, or a protein			
CC	that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or			
CC	identical to an entire amino acid sequence of PI; (2) a polynucleotide			
CC	that encodes the protein; (3) a composition comprising a polynucleotide			
CC	that is a full complement of the polynucleotide described above; (4)			
CC	generating a mammalian immune response directed to the protein of PI; (5)			
CC	detecting, in a sample, the presence of a 254PID6B-related protein or a			
CC	254PID6B-related polynucleotide; (6) a composition that modulates the			
CC	status of a cell that expresses a protein of PI; (7) a pharmaceutical			
CC	composition that comprises the composition described above in a human			
CC	unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)			
CC	a non-human transgenic animal that produces the antibody; (10) a			
CC	hybridoma that produces the antibody; (11) delivering a cytotoxic agent			
CC	or a diagnostic agent to a cell that expresses the protein of PI; and			
CC	(12) inhibiting growth, reproduction or survival of cancer cells that			
CC	express the protein of PI. 254PID6B has cytostatic activity, and can be			
CC	used in gene therapy. The compositions, molecules and methods are useful			
CC	for treating and detecting cancer. The present sequence represents the			
CC	human 254PID6B v.1 nucleotide sequence, which is used in the			
CC	exemplification of the present invention. The human 254PID6B gene is			
CC	located on chromosome 6p22.			
XX				
SQ	Sequence 6797 BP; 1812 A; 1542 C; 1687 G; 1756 T; 0 U; 0 Other;			
Alignment Scores:				
Pred. No.:	0	Length:	6797	

Qy	1	MetAlaProThrGlyValLeuSerSerLeuLeuLeuValThrIleAla---	G1	19
Db	514	ATGGGCCCCCAGAGTGTCTCTTCATTTGCTGCTGGTGGACAAATTCAGTTGG		573
Qy	19	YCysAlaArgLysGlnCysSerGluGlyArgThrYzSerAsnAlaValIleSerProAs		39
Db	574	TTGTGCCGTAAGCAGTGCAGCGAGGGGAGACATATTTCCAATGCAGTCAITTCACCTAA		633
Qy	39	nLeuGluThrThrArgIleMetArgValSerHisThrPheProValValAspCysThrAl		59
Db	634	CTTGGAACACCAACAGATCATGCGGTGTCTCACACCTTCCCTGCTAGACTGCACGCGC		693
Qy	59	aAlaCysCysAspLeuSerSerCysAspLeuAlaTTPTrPheGluGlyArgCysTyLe		79
Db	694	CGTTTGTCTGTGACCTGTCCAGCTGTGACCTGGCTGTGGTTCGAGGGCCGCTGTACT		753
Qy	79	uValSerCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTy		99
Db	754	GGTGAGCTGCCCCCACAAAGAGAACTGTGAGCCCAAGAGATGGGCCCATCAGGTCTTA		813
Qy	99	rLeuThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyGlyAspMe		119
Db	814	TCTCAGCTTTTGTCTCGGGCTGTTCAGAGCCCTGCACAGCTGTCTGGACTATGGGACAT		873
Qy	119	tMetLeuAsnArgGlySerProSerGlyIleTTPTrGlyAspSerProGluAspIleArgLy		139
Db	874	GATCTGAACAGGGGCTCCCCCTCGGGATCTGGGGGACCTCAGCTGAGGATATCAGAAA		933
Qy	139	sAspLeu***PheLeuGlyLysAspTrpGlyLeuGluGluMetSerGluTy***AspAs		159
Db	934	GGACTTGCCCTTTCTAGGCAAGATTTGGGGCTAGAGGAGATGTCTGAGTACTCAGATCA		993
Qy	159	pTyArgGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySe		179
Db	994	CTACCGGAGCTGGAGAGAGGACCTCTTGCAACCCAGTGGCAAGCAGAGAGCCAGAGGAG		1053
Qy	179	rAlaGluTyThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerVa		199
Db	1054	TGCCAGGTACAGGACTGGGGCTTACTCCGGGAGCAGAGGGGGCTTCAACTCTCTGT		1113
Qy	199	lGlyAspSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyLeuAs		219
Db	1114	TGGAGACAGTCTCGGGTGCCAGCGGAGACGACGAGGAGCCCTGAGCTCCATTACTGAA		1173
Qy	219	nGluSerAlaSerThrProAlaProLysLeuProLysArgSerValLeuLeuProLeuPr		239
Db	1174	TGAGTCGGCTTCAACCCCTGCCCCAAAACCTCCCTGAGAGAAAGTGTGTCTTCCCTTGC		1233
Qy	239	oThrThrProSerSerGlyValLeuGluLysGluLysAlaSerGlnLeuGlnGluG1		259
Db	1234	GACTACTCAITCTTCAGAGAGGTGTTCGAGAAAGAAAGGCTTTCAGTTCAGGAAACA		1293
Qy	259	nSerSerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProAlaSe		279
Db	1294	ATCCAGCAACAGCTCTGGAAAAGAGGTTCTAATGCTTCCCATAGTCTTCCCTCCGCAAG		1353
Qy	279	rLeuGluLeuSerSerValThrValGluLysSerProValLeuThrValThrProGlySe		299
Db	1354	CCTGGAGCTCAGCTCAGTCACCGTGAGAAAAGCCAGTGCAGTCAACCCCGGGAG		1413
Qy	299	rThrGluHisSerIleProThrProThrSerAlaAlaProSerGluSerThrProSe		319
Db	1414	TACAGACACAGCATCCCAACACCTCCCACTAGCAGAGCCCCCTCTGAGTCCACCCATC		1473
Qy	319	rgLeuLeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAl		339

Db 1474 TGAAGTACCCATATCTCTTACCACTGCTCCAGACAGTGAAGAACTTACGGTATCGGC 1533  
Qy 339 aGlyAspAsnLeuIlelleThrLeuProAspAsnGluValGluLeuIleValAlaPheValAl 359  
Db 1534 TGGAGATAACCTAATTAATCTTACCCGCAATGAAGTTGAACTGAAGGCGCTTTGTTGC 1593  
Qy 359 aProAlaProValGluThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 379  
Db 1594 GCCAGGCCACCTGTAGAAACAACCTACAACTATGAATGGAATTTAAATAGCCACCCAC 1653  
Qy 379 rAspTyrGlnGlyGluIleValGlyHisGlnThrLeuAsnLeuSerGlnLeuSe 399  
Db 1654 AGACTACCAAGTGAATTAACCAAGGACACCAAGCAACTCTTAACTCTCTCAATTGTC 1713  
Qy 399 rValGlyLeuTyrValPheIleValThrValSerSerGluAsnAlaPheGlyGluGlyPh 419  
Db 1714 CGTCGGACTTATGCTCTCAAGTCACTGTTCTTAGTGAAACGCCCTTTGGAGAAGGATT 1773  
Qy 419 eValAsnValThrValIleValProAlaArgValAsnLeuProProValAlaValSe 439  
Db 1774 TGTCAATGTCACTTTAAGCGCTGCCAGAAGAGTCAACCTGCCACCTGTAGCAGTTGTTTC 1833  
Qy 439 rProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSe 459  
Db 1834 TCCCAACTGCAGAGCTCACTTGGCTTTGACGTCAGCCCTCAATTGATGGCACCCAAAG 1893  
Qy 459 rThrAspAspThrGluIleValSerTyrHisTrpGluGluIleAsnGlyProPheIleGl 479  
Db 1894 TACAGATGATACTGAATAGTAGTATTATCAATTGGGAAGAAATAAACGGCCCTTCATAGA 1953  
Qy 479 uGluIleThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTy 499  
Db 1954 AGAAGAAGCTTCAGTTGACTCTCCGCTTACGCTTGTCTAAACCTTGATCCGTGTAACCTA 2013  
Qy 499 rSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLe 519  
Db 2014 TAGTTTCAGGTGACTGTTACAGACTCGACGGAGCCACTAACTCTCAACTGCGAGCCCT 2073  
Qy 519 uIleValAsnAsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIl 539  
Db 2074 AATAGTGAACAATGCTGTGGACTACCCACCAGTTGCTAATGCAGGACCAATCACACCAT 2133  
Qy 539 eThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIl 559  
Db 2134 AACTTTGCCCCAAACCTCCATCACTTTGAATGGAAACACAGACGAGTCAACCATCACAGAT 2193  
Qy 559 eValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyIleHisValValMetGlnI 579  
Db 2194 TGTCTCTATAGTGGTCCCTGGGTCTCTGGAGTGAGGGGCAAAACATGTGGTCAATGCAGGG 2253  
Qy 579 yValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLe 599  
Db 2254 AGTACAGCGCATACCTTCATTTATCTGCAATGCAGGAGGAGATTATACATTTTCAGCT 2313  
Qy 599 uLysValThrAspSerArgGlnGlnSerThrAla\*\*\*ValThrValIleValGlnPr 619  
Db 2314 GAAGGTGACAGATTCTTCAGGCCAACAGTCTACTGCTGTGGTGTGCTGTGATGTGCCAGCC 2373  
Qy 619 oGluAsnAsnArgProProValAlaValAlaGlyProAspIleValLeuIlePheProVa 639  
Db 2374 TGAATAACAATAGACCTCCAGTGGCTGTGGCGGCCCTGTATAAAGAGCTGATCTTCCAGT 2433  
Qy 639 lGluSerAlaThrLeuAspGlySerSerSerAspAspHisGlyIleValPheTyrHi 659  
Db 2434 GGAAGTGTACCTCTGGATGGAGACGACGACGAGTGACCCAGCATTTGCTCTTACCA 2493  
Qy 659 sTrpGluHisValArgGlyProSerAlaValGluMetGluAsnIleAspIleAlaI 679  
Db 2494 CTGGGAGCACGTACAGAGGCCCCAGTGCAGTGGAGATGGAAATATTGACAAAGCAATAGC 2553  
Qy 679 aThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrValIleAspGlnGl 699

Db 2554 CACTGTGACTGGTCTCCAGGTGGGACCTACCACCTTCGGTTTGCACGTGAAGACCAGCA 2613  
Qy 699 nGlyLeuSerSerThrSerThrLeuThrValAlaValIleGlySerGluAsnSerProPr 719  
Db 2614 GGGACTGAGCAGCAGCTCCACCTCACTGTGGCTGTGAAGAGGAAAAATAATAGTCTCTCC 2673  
Qy 719 oArgAlaArgAlaGlyGlyArgHisValLeuLeuLeuProAsnAsnSerIleThrLeuAs 739  
Db 2674 CAGAGCCCGGCTGGTGGCAGACATGTTCTTGTGCTTCCCAATAATTCATATTCTTGGGA 2733  
Qy 739 pGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuTrpIleArgAspGlyGl 759  
Db 2734 TGGTTCAAGTCTACTGATGACCAAGAAATTTGTGCTCTATCTGTGGATCCGGGATGGCA 2793  
Qy 759 nSerProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuTh 779  
Db 2794 GAGTCCACGACGCTGGAGATGTCATCGATGGCTCTGCACACAGTGTGGCTCTGCAGCTTAC 2853  
Qy 779 rAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSe 799  
Db 2854 GAATCTGGTGGAGGGGTGTACACTTTCCACTTTCGAGTCAACCACAGTCAAGGGGGCCTC 2913  
Qy 799 rAspThrAspThrAlaThrValGluValGlnProAspProArgIleValSerGlyLeuValGl 819  
Db 2914 GGACACAGACACTGCCACTCTGTGAGGTGCAGCCACAGACCTTAGGAAGAGTGGCCTTGGTGA 2973  
Qy 819 uLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgIleValSerThrLeuValAr 839  
Db 2974 GCTGACCTCGAGGTTGGTGTGGGCACTGCACAGACGACGCGAAGGACACCCCTTGTGAG 3033  
Qy 839 gGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleValGlnIleValGlnIleValArgAl 859  
Db 3034 GCAGCTGGCTGTGCTGTGACCGTGTGAGCTCGACATTAAGGTCCAGAAGATTCGGGC 3093  
Qy 859 aHisSerAspLeuSerThrValIleValPheTyrValGlnSerArgProProPheLysVa 879  
Db 3094 CCACTCGATCTCAGCACCGTGTGTTTGTATTGTATACAGACGACGCGCCTTTCAGGT 3153  
Qy 879 lLeuLysAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerIleValGlyAlaAs 899  
Db 3154 TCTCAAGCTGCTGAAGTGGCCGAAATCTGCACATCGGCTCTCAAAAGGAGAAGGTGA 3213  
Qy 899 pPheLeuLeuPheIleValLeuArgValAspThrAlaGlyCysLeuLeuIleValCysSerGl 919  
Db 3214 CTTCTTCTCTTTCAGGCTTGTGAGGTTGTATACAGCAGGTGGCCTTCTGAAGTGTCTGG 3273  
Qy 919 yHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAs 939  
Db 3274 CCATGTCATCGACCCCTCTCAAAAGCGCTGCATTTGCTCTCCTATATGGATGGAGAA 3333  
Qy 939 nLeuIleGlnArgTyrIleTrpAspGlyGluSerAsnCysGluTrpSerIlePheTyrVa 959  
Db 3334 CTTTATACAGCTTATATCTGGGATGAGAGAGCAACTGTGAGTGGAGTATATCTATGT 3393  
Qy 959 lThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCy 979  
Db 3394 GACAGTGTGGCTTTACTCTTATTTGCTTAACAGGAGGTTTCACTTGGCTTTCGACTCTG 3453  
Qy 979 sCysCysLysArgGlnLysArgThrLysIleAlaGlyLysThrLysTyrThrIleLeuAs 999  
Db 3454 CTGCTGCAAAAGACAAAAAGGACTAAATCAGGAAAAAACAAGTACACCATCTCTGGA 3513  
Qy 999 pAsnMetAspGluGlnGluArgMetGluLeuArgProLysTyrGlyIleLysHisArgSe 1019  
Db 3514 TAACATGATGAACAGAGAAAGATGGAACCTGAGSCCCCAATAATGTTATCAAGCACCCGAG 3573  
Qy 1019 rThrGluHiAsnSerSerLeuMetValSerGluSerGluPheAspSerAspGlnAspTh 1039  
Db 3574 CACAGACACAACCTCCAGCCTGTATGTTATCCGAGTCTCGAGTTCGACAGTCCACGAGCAC 3633  
Qy 1039 rIlePheSerArgGluLysMetGluArgGlyAsnProLysValSerMetAsnGlySerIl 1059  
Db 3634 AATCTTCAGCCGAGAAAGATGGAGAGGGGAATCCAAAGGTTTCCATGAATGTTTCCAT 3693

QY 1059 eArgAsnGlyAlaSerPheSerTyrCysSerLysAspArg 1072  
 |||||  
 Db 3694 CAGAAATGGAGCTTCCTTCAGTTATGCTCAAGGACAGA 3733

RESULT 5

ADRO0595  
 ID ADRO0595 standard; cDNA; 6991 BP.  
 XX  
 AC ADRO0595;  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE Human 254PID6B v.3 encoding cDNA SEQ ID NO:6.  
 XX  
 KW 254PID6B; small interfering RNA; siRNA; immune response;  
 KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254PID6B v.3; gene; ss; chromosome 6.  
 XX  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH CDS 739..3930  
 FT /\*tag= a  
 FT /product= "254PID6B v.3"

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.  
 DR P-PSDB; ADRO0596.

XX  
 PT New 254PID6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

PS Claim 10; SEQ ID NO 6; 345pp; English.

XX  
 CC The present invention describes a 254PID6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID  
 CC NO:3, 5 or 7 ADRO0592, ADRO0594 or ADRO0596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of PI; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of PI; (5)  
 CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
 CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of PI; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of PI; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that

CC express the protein of P1. 254PID6B has cytoskeletal activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence encodes the human  
 CC 254PID6B v.3, which is used in the exemplification of the present  
 CC invention. The human 254PID6B gene is located on chromosome 6p22.

XX  
 SQ Sequence 6991 BP; 1873 A; 1589 C; 1738 G; 1791 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 6991  
 Score: 5527.00 Matches: 1068  
 Percent Similarity: 96.83% Conservatives: 0  
 Best Local Similarity: 96.83% Mismatches: 4  
 Query Match: 99.05% Indels: 31  
 DB: 13 Gaps: 1

US-10-764-390-3 (1-1072) x ADRO0595 (1-6991)

QY 1 MetAlaProThrGlyValLeuSerSerLeuLeuLeuValThrIleAla----- 18  
 |||||  
 Db 621 ATGGGCCCCCCCCACAGGTGTCTCTCTTCAATTGCTGCTGGTGACAAATTGCAGTTTGC 680  
 |||||  
 QY 18 ----- 18  
 |||||  
 Db 681 TTATGTTGATGTCATCATGGCAAAAAAATCATCTGTTGAGCATCATTTAAGAAGACCCAT 740  
 |||||  
 QY 19 -----GlyCysAlaArgLysGlnCysSerGluGlyArgTh 30  
 |||||  
 Db 741 GACTAGACTGGCTGGCGGAGCCCATGTTGTGCCGTAAAGCAGTGCGAGGAGGAGGAC 800  
 |||||  
 QY 30 rTySerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSerHi 50  
 |||||  
 Db 801 ATATTCCAATGCAGTCATTTCACCTTACTTGGAAACACACAGAAATCATGCGGTGTCTCA 860  
 |||||  
 QY 50 sThrPheProValValAspCysThrAlaAlaCysCysAspLeuSerSerCysAspLeuAl 70  
 |||||  
 Db 861 CACCTTCCTCTGCTAGACTGACGGCGCTTGTGTGACCTGTCTCAGCTGTGACCTGGC 920  
 |||||  
 QY 70 aTTPTrPheGluGlyArgCysTyrLeuValSerCysProHisLysGluAsnCysGluPr 90  
 |||||  
 Db 921 CTGGTGGTTCAGGGCGCTGCTACCTGGTGAGCTGCCCGCCCAACAAAGAGAACTGTGAGCC 980  
 |||||  
 QY 90 oLysLysMetGlyProIleArgSerTyrLeuThrPheValLeuArgProValGlnArgPr 110  
 |||||  
 Db 981 CAAGAAGATGGCCCCCATCAGTCTTATCTCACTTTTGTCTCGCGCTGTTTCAGAGGCC 1040  
 |||||  
 QY 110 oAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgLysSerProSerGlyIleTr 130  
 |||||  
 Db 1041 TGCAAGCTGCTGGACTATGGGACATGATGCTGAACAGGGGCTCCCTCTGGGGATCTG 1100  
 |||||  
 QY 130 pGlyAspSerProGluAspIleArgLysAspLeu\*\*\*PheLeuGlyLysAspTIPGlyLe 150  
 |||||  
 Db 1101 GGGGACTCAGCTGAGGATATCAGAAAGACCTTGCCCTTTCTAGGCAAGATTTGGGGCT 1160  
 |||||  
 QY 150 uGluGluMetSerGluTyr\*\*\*AspAspTyrArgGluLeuGluLysAspLeuLeuGlnPr 170  
 |||||  
 Db 1161 AGAGGAGATGCTGAGTACTCAGATGACTACCGGGAGCTGGAGAAGGACCTCTTCAACC 1220  
 |||||  
 QY 170 oSerGlyLysGlnGluProArgGlySerAlaGluTyrThrAspTyrGlyLeuLeuProGl 190  
 |||||  
 Db 1221 CAGTGGCAAGCAGGAGGCCAGAGGAGTGGCGAGTACCGGACTGGGGCTTACTGCGGG 1280  
 |||||  
 QY 190 ySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThrGl 210  
 |||||  
 Db 1281 CAGCAGGGGGCTTCAACTCTCTTTGGAGACAGTCTCTGGGTGCCAGCGGAGACGCA 1340  
 |||||  
 QY 210 mGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeuPr 230  
 |||||  
 Db 1341 GCAGGACCTGAGCTCCATTACTTGAAATGAGTGGCTTCAACCTCTGCCCAAACTCCC 1400  
 |||||  
 QY 230 oGluArgSerValLeuLeuProLeuProThrThrProSerSerGlyGluValGluGly 250  
 |||||  
 Db 1401 TGAGAGAAATGTTGTTGCTTCCCTTCCCGACTACTCTCACTTTCAGGAGAGGTGTGGAGAA 1460  
 |||||

Qy 250 sGluIysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGlyIysGluValLeuMe 270  
Db 1461 AGAAAGAGGCTTCTCAGCTCCAGGAAACAATCCAGCAACAGCTCTGGAAAGAGGTTCTAAT 1520  
Qy 270 tProSerHisSerLeuProProAlaSerLeuGluLeuSerSerValThrValGluIysSe 290  
Db 1521 GCCTTCCATAGTCTTCTCCGGCAGCCTGGAGCTCAGCTCAGTCACCGTGGAGAAAG 1580  
Qy 290 rProValLeuThrValThrProGlySerThrGluHisSerIleProThrProThrSe 310  
Db 1581 CCCAGTGCTCACAGTCACCCCGGGAGTACAGAGCACAGCATCCCAACACCTCCCACTAG 1640  
Qy 310 rAlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaProAr 330  
Db 1641 CGCAGCCCTCTGAGTCCACCCCATCTGAGCTACCCATATCTCTACCACTGCTCCAG 1700  
Qy 330 gThrValIysGluLeuThrValSerAlaGlyAsnLeuIleIleThrLeuProAspAs 350  
Db 1701 GACAGTGAAGAACTTACGGTATCGCTGGAGATAACCTAATATTAACCTTTACCCGACAA 1760  
Qy 350 nGluValGluLeuIysAlaPheValAlaProAlaProValGluThrThrIyrAsnTy 370  
Db 1761 TGAAGTTCAACTGGAAGGCTTTGTTGCGCCAGCGCACCTGTAGAAAACAACCTACAACCTA 1820  
Qy 370 rGluTrpAsnLeuIleSerHisProThrAspTyrgGlnGlyIleIysGlnGlyHisIy 390  
Db 1821 TGAATGGAAATTAATAAGCCACCCACAGACTACCAAGGTGAATAAACAAGGACACAA 1880  
Qy 390 sGlnThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrrValPheIysValThrValSe 410  
Db 1881 GCRAACTCTTAACCTCTCTCAATGTGCTGGACTTTATGTCTTCAANGTCACGTGTTTC 1940  
Qy 410 rSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValIysProAlaArgVa 430  
Db 1941 TAGTGAAAGCCCTTTGAGAGGATTTGTCAATGTCACTGTAAAGCTGCCAGAGAGT 2000  
Qy 430 lAsnLeuProProValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuth 450  
Db 2001 CAACCTGCACCTGTAGCAGTGTGTTCTCCCAACTGCAGAGCTCACTTTGCCCTTTGAC 2060  
Qy 450 rSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSerTyrrHisTr 470  
Db 2061 GTCAGCCCTCAATGATGCGACGCAAAAGTACAGATGATACTGAAATAGTATGATTCATTG 2120  
Qy 470 pGluGluIleAsnGlyProPheIleGluGluIysThrSerValaAspSerProValLeuAr 490  
Db 2121 GGAAGAAATAAACGGGCCCTTCATAGAGAGAGAGACTTCAGTTGACTCTCCCGTCTTACG 2180  
Qy 490 gLeuSerAsnLeuAspProGlyAsnTyrrSerPheArgLeuThrValThrAspSerAspGl 510  
Db 2181 CTTGTCTAACCTTGATCTGTGTAATACTATAGTTTCAGGTTGACTGTTCAGACTCGGACGG 2240  
Qy 510 yAlaThrAsnSerThrAlaAlaLeuIleValAsnAsnAlaValaAspTyrrProProVa 530  
Db 2241 AGCCACTAACCTCTACACTGACGCCCTTAATAGTGAACAATGCTGGACTACCCACCACT 2300  
Qy 530 lAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGl 550  
Db 2301 TGCTAATGACGAGCAACATCACACATTAACCTTTGCCCCAAAACCTCCATCACTTTGAATGG 2360  
Qy 550 yAsnGlnSerSerAspAspHisGlnIleValLeuTyrrGluTrpSerLeuGlyProGlySe 570  
Db 2361 AAACAGAGCAGTGACGATCACAGATGTCCTCTATGAGTGGTCCCTGGGCTCTGGAG 2420  
Qy 570 rGluGlyIysHisValValMetGlnGlyValGlnThrProTyrrLeuHisLeuSerAlaMe 590  
Db 2421 TGAGGGCAAAACATGCGGTATGTCAGGGAGTACAGACGCCATACCTTCATTATCTGCAAT 2480  
Qy 590 tGlnGluGlyAspTyrrThrPheGlnLeuIysValThrAspSerSerArgGlnGlnSerTh 610  
Db 2481 GCAGGAGAGATTAATACATTTACGCTGAAGGTGACAGATTCTTCAAGGCAACAGTCTAC 2540

Qy 610 rAla\*\*\*ValThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaGl 630  
Db 2541 TGCTGTGCTGACTGTGATTTGTCCAGCCTGAAAACAATAGACCTCCAGTGGCTGTGGCCGG 2600  
Qy 630 yProAspIysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSe 650  
Db 2601 CCCTGATAAAGAGCTGATCTTCCAGTGGAAAGTGTACTCCCTGGATGGAGCAGCAGCAG 2660  
Qy 650 rAspAspHisGlyIleValPheTyrrHisTrpGluHisValArgGlyProSerAlaValGl 670  
Db 2661 CGATGACCACCGCATTTGCTTACCACTGGGAGCAGCTCAGAGGCCCCAGTCGACGTGGA 2720  
Qy 670 uMetGluAsnIleAspIysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrrHi 690  
Db 2721 GATGGAAATATTTGACAAAGCAATAGCCACTGTGACTGTCTCCAGGTGGGAGCTTACCA 2780  
Qy 690 sPheArgLeuThrValIysAspGlnGlnIysSerSerThrSerThrLeuThrValAl 710  
Db 2781 CTTCCGTTTGACAGTGAAGACACAGCAGGACTCAGCAGCAGCTCCACCTCCTGCTGCGC 2840  
Qy 710 aValIysGluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuVa 730  
Db 2841 TGTGAAGAAGGAAATAATAGTCTCCAGAGCCCGGCTGGTGGCAGACATGTTCTTGT 2900  
Qy 730 lLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleVa 750  
Db 2901 GCTTCCCAATAATTTCCATTTACTTTGGATGGTTCAAGGTCTACTGATCCAAAGAAATGT 2960  
Qy 750 lSerTyrrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySe 770  
Db 2961 GTCTATCTGTGGATCCGGGATGCCAGAGTCCAGAGCTGAGATGTCTATCGATGGCTC 3020  
Qy 770 rAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrrPheHisIle 790  
Db 3021 TGACCACAGTGTGCTCTGCAGCTTACGAATCTGGTGGAGGGGTGTACACTTTCCACTT 3080  
Qy 790 uArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnPr 810  
Db 3081 GCGAGTCAACCACAGTCAAGGGGCTCGGACACAGACACTGCCACTGTGGAGTGCAGCC 3140  
Qy 810 oAspProArgIysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuTh 830  
Db 3141 AGACCTTAGGAGAGTGGCTGTGGAGCTGACCTGCAGGTTGGTGTGGCAGCTGAC 3200  
Qy 830 rGluGlnArgIysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSe 850  
Db 3201 AGACGACGGGAAGACACCTTGTGAGCAGCTGGCTGTGCTGTCTGAAACGTGTGAGCTC 3260  
Qy 850 rAspIleIysValGlnIysIleArgAlaHisSerAspLeuSerThrValIleValPheTy 870  
Db 3261 GGACATTAAGGTCAGAGAGATTCGGGCCCACTCGGATCTCAGACCCGTTGTTGTTTTA 3320  
Qy 870 rValGlnSerArgProPheIysValLeuIysAlaAlaGluValAlaArgAsnLeuHi 890  
Db 3321 TGTCAGACGACGGCCGCTTTCAAGGTTCTCAAGCTGTCTGAAGTGGCCCCGAAATCTGCA 3380  
Qy 890 sMetArgLeuSerIysGluIysAlaAspPheLeuPheIysValLeuArgValAspTh 910  
Db 3381 CATCGGCTCTCAAGGAGAGAGGCTGACTTCTGCTTTTCAAGTCTTGGGGTTGATAC 3440  
Qy 910 rAlaGlyCysLeuLeuIysCysSerGlyHisGlyHisCysAspProLeuThrIysArgCy 930  
Db 3441 AGCAGGTTGCCCTTCTGAAGTGTCTGGCCATGGTCACTCGCACCCCTTCAAAAAGCGCTG 3500  
Qy 930 sIleCysSerHisLeuTrpMetGluAsnLeuIleGlnArgTyrrIleTrpAspGlyVGluse 950  
Db 3501 CATTTGCTCTCACTTATGATGGAGAACTTATACGGTTTATATCTGGGATGGAGAGAG 3560  
Qy 950 rAsnCysGluTrpSerIlePheTyrrValThrValLeuAlaPheThrLeuIleValLeuTh 970  
Db 3561 CAACGTGAGTGGAGTATATTTCTATGTGACAGTGTGGCTTTTACTTTTACTTTTGTGCTAAC 3620  
Qy 970 rGlyGlyPheThrTrpLeuCysIleCysCysLeuArgGlnIysArgThrIysIleAr 990







Db 1221 CAGTGGCAAGCAGGAGCCAGGAGGAGTCCGAGGTACAGGACTGGGCGCTACTGGCCGG 1280  
Qy 190 ySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThrGly 210  
Db 1281 CAGCAGGGGGCCCTTCAACTCTCTGTTGGAGACAGCTCTCGGCTGCCAGCGGAGCGCA 1340  
Qy 210 nGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeuPr 230  
Db 1341 CGAGGACCTGAGCTCCATACCTGMAATGAGTGGCTTCAACCCCTGCCCAAACTCCC 1400  
Qy 230 oGluArpSerValLeuLeuProLeuProThrThrProSerSerGlyGluValLeuGluLy 250  
Db 1401 TGAGAGAAGTGTGTGCTTCTCCCTGCGCACTCTCATCTTCAGGAGAGGTGTGGAGAA 1460  
Qy 250 sGluLysAlaSerGlnLeuGlnSerSerAsnSerSerGlyLysGluValLeuMe 270  
Db 1461 AGAAAGGCTTCTCAGCTCCAGGACCAATCCAGCAACAGCTCTCGAAAGAGGTTCTAAT 1520  
Qy 270 tProSerHisSerLeuProAlaSerLeuGluLeuSerSerValThrValGluLysSe 290  
Db 1521 GCCTTCCCATAGTCTTCTCCGGCAAGCTCGAGCTCAGCTCAGTCCAGCTGGGAGAAAG 1580  
Qy 290 rProValLeuThrValThrProGlySerThrGluHisSerIleProThrProThrSe 310  
Db 1581 CCCAGTGTCTCAGCTCACCCCGGGAGTACAGAGCACAGCATCCCAACACTCCCACTAG 1640  
Qy 310 rAlaAlaProSerGluSerThrProSerGluLeuProIleSerProThrThrAlaProAr 330  
Db 1641 CGCAGCCCTCTCAGTGTACCCCATCTGAGCTACCCATATCTCTACCACTGTCTCCAG 1700  
Qy 330 qThrValLysGluLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAs 350  
Db 1701 GACAGTGAAGAACTACGGTATCGGCTGGAGATAACCTTAATTAATCTTTACCCGCAAA 1760  
Qy 350 nGluValGluLeuLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTy 370  
Db 1761 TGAAGTTGAACCTGAAGGCTTGTGTGCCAGCGCCACCTGTAGAAACAACCTACACTA 1820  
Qy 370 rGluTrpAsnLeuIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLy 390  
Db 1821 TGAATGGAAATTAATAAGCCACCCACAGACTACCAAGGTGAATTAATAAACAAGACAAA 1880  
Qy 390 sGluThrLeuAsnLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSe 410  
Db 1881 GCAAACTCTTAACCTCTCTCAATTTGTCCGTCCGACTTATATCTCTCAAGTCACTGTTTC 1940  
Qy 410 rSerGluAsnAlaPheGlyGluGlyPheValAsnValThrValLysProAlaArgArgVa 430  
Db 1941 TAGTGAATAACCCCTTGGAGAGGATTTGTCAATGTCACTGTTAAGCTGCCAGAGAGT 2000  
Qy 430 lAsnLeuProProValAlaValSerProGlnLeuGlnGluLeuThrLeuProLeuTh 450  
Db 2001 CAACCTGCCACTGTAGCAGTGTGTTCTCCCAACTGCAAGAGCTCACTTTGCCCTTTGAC 2060  
Qy 450 rSerAlaLeuIleAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTr 470  
Db 2061 GTACGCCCTCATGTAGCCAGCCCAAGTACAGATGATACCTGAATAGTGTATCATTTG 2120  
Qy 470 pGluGluIleAsnGlyProPheIleGluGluLysThrSerValaAspSerProValLeuAr 490  
Db 2121 GGAAGAAATAAACCGGCCCTTCATAGAGAGAGACATTCAGTTGACTCTCCCGCTTACG 2180  
Qy 490 gLeuSerAsnLeuAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGl 510  
Db 2181 CTTGTCTAACCTTGATCTCTGTTAACTATAGTTTCAGGTTGACTGTACAGACTCGGACGG 2240  
Qy 510 yAlaThrAsnSerThrThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProProVa 530  
Db 2241 AGCCACTAACTCTCAACTGACGCCCTTAATAGTGAACAAATGCTGTGGACTACCCACCACT 2300  
Qy 530 lAlaAsnAlaGlyProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGl 550  
Db 2301 TGTAAATGCGGACCAAAATCAACATAAATTTGCCCAAAACTCCATCACTTTTGAATGG 2360

Qy 550 yAsnGlnSerSerAspAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySe 570  
Db 2361 AAACACAGAGAGTACGATCACCAATTGTCTCTATGAGTGGTCCCTGGGTCTCTGGAG 2420  
Qy 570 xGluGlyLysHisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMe 590  
Db 2421 TGAGGGCAAAATGTGTGTATGAGGAGGTACAGCGCCATACCTTCATTTATCTGCANT 2480  
Qy 590 tGlnGluGlyAspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerTh 610  
Db 2481 GCAGAAAGGAGATTATACATTTAGCTGAAGTGCACAGATTCTTCAAGGCAACAGTCTAC 2540  
Qy 610 rAlaA\*\*ValThrValIleValGlnProGluLysAsnAsnArgProProValAlaValAlaGl 630  
Db 2541 TGCTGTGGTACTGTGATTGTCCAGCTCGAATAAATAAGACCTCCAGTGGCTGTGGCCGG 2600  
Qy 630 yProAspLysGluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSe 650  
Db 2601 CCCTGATTAAGAGCTGATCTTCCAGTGGAAAGTCTTACCTCGATGGGACGACGAGCG 2660  
Qy 650 rAspAspHisGlyIleValPheTyrHisTrpGluHisValArgGlyProSerAlaValGl 670  
Db 2661 CGATGACACCGCAATTGTCTTACCACTGGGACACGTCAGAGGCCCCAGTGCAGTGA 2720  
Qy 670 uMetGluAsnIleAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHi 690  
Db 2721 GATGAAATAATTATGACAAAGCAATAGCCACTGTGACTGGTCTCCAGSTGGGACCTACCA 2780  
Qy 690 sPheArgLeuThrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAl 710  
Db 2781 CTTCCGTTTGACAGTGAAGACGACGAGGACTGAGCAGCACGTCACCCCTCACTGTGGC 2840  
Qy 710 aValLysGlyGluAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuVa 730  
Db 2841 TGTGAAGAGGAAATAATAGTCTCTCCAGAGCCCGGCTGGTGGCAGACATGTTCTTGT 2900  
Qy 730 lLeuProAsnAsnSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleVa 750  
Db 2901 GCTTCCCAATAATTTCCATTACTTTTGGATGGTTCAAGGTCTACTGATGACCAAGAAATGT 2960  
Qy 750 lSerTyrLeuTrpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySe 770  
Db 2961 GTCTATCTGTGGATCCGGATGGCCAGAGTCCAGCAGCTGGAGATGTCTCATGATGGCTC 3020  
Qy 770 rAspHisSerValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLe 790  
Db 3021 TGACCACAGTGTGGCTCTGCAGCTTACGAATCTGGTGGAGGGGTGTACACTTTCCACTT 3080  
Qy 790 uArgValThrAspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnPr 810  
Db 3081 GCGAGTCACCGACAGTCAAGGGGCTCCGACACAGACACTGCCACTGTGGAAAGTGCAGCC 3140  
Qy 810 oAspProArgLysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuTh 830  
Db 3141 AGACCTTAGAAGAGTGGCTGTGGAGCTGACCCCTCGAGGTGGTGTGGGACGTGAC 3200  
Qy 830 rGluGlnArgLysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSe 850  
Db 3201 AGACGACGGNAGGACACCTTGTGAGGACAGCTGGCTGTGCTGTGAACGTGTGGACTC 3260  
Qy 850 rAspIleLysValGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTy 870  
Db 3261 GGACATTAAGGTCAGAAAGATTGGGGCCACTCGGATCTCAGCACCGCTGATGTGTGTTTA 3320  
Qy 870 rValGlnSerArgProProPheLysValLeuLysAlaGluValAlaArgAsnLeuHi 890  
Db 3321 TGTACAGAGACGGCCCTTTCAAGGTTCTCAAAGCTGCTGAAGTGGCCCAAACTGCA 3380  
Qy 890 sMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuArgValAspTh 910  
Db 3381 CATGCGGCTCTCAAGAGAGAGGCTGACTTCTTGTCTTTTCAAGGCTCTTGAGGGTTGATAC 3440



101 ThrPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMet 120  
120 Db  
367 ACTTTTGTGCTCCGGCCCTGTTTCAGAGGCTTCACAGCTGCTGGACTATGGGGAATGATG 426  
121 LeuAsnArgGlySerProSerGlyIleTyrGlyAspSerProGluAspIleArgGlyAsp 140  
140 Db  
427 CTGAACAGGGGCTCCCTCGGGGATCTGGGGGACTCACCTGAGGATATCAGAAAGGAC 486  
141 Leu\*\*\*PheLeuGlyLysAspTyrGlyLeuGluMetSerGluTyr\*\*\*AspAspTyr 160  
160 Qy  
487 TTGCCCTTCTTAGGCAAAAGATTGGGGCTAGAGGAGATGTTGAGTACTCAGATGACTAC 546  
161 ArgGluLeuGlyLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAla 180  
180 Qy  
547 CGGGAGCTGGAGAGGAGCTCTTGCACCCAGTGGCAAGCAGGAGGCCAGAGGGAGTGCC 606  
181 GluTyrThrAspTyrGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGly 200  
200 Qy  
607 GAGTACACGAGCTGGGGCTACTGCCGGGACGAGGAGGAGGCTTCACTCTCTCTGTTGA 666  
201 AspSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGlu 220  
220 Qy  
667 GACAGCTCTGGGTGCCAGCGAGAGCAGCAGCAGGAGCCCTGAGCTCCATTCACTCTGAATGAG 726  
221 SerAlaSerThrProAlaProLysLeuProGluArgSerValLeuLeuProLeuProThr 240  
240 Qy  
727 TCGGCTTCAACCCCTGCCCCAAAACCTCCCTCAGAGAGTGTGTGCTTCCCTTCCCGACT 786  
241 ThrProSerSerGlyGluValLeuGluLysGlyLysAlaSerGlnLeuGlnGlnSer 260  
260 Qy  
787 ACTCACTTTCAGAGAGGTGTGGAGAAAGAAAGAGGTCTTCAGCTCCAGAACATCC 846  
261 SerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProProAlaSerLeu 280  
280 Qy  
847 AGCAACAGCTCTGGAAAGAGGTCTTAATGCTTCCCATAGTCTTCTCTCCCGCAAGCCTG 906  
281 GluLeuSerSerValThrValGluLysSerProValLeuThrValThrProGlySerThr 300  
300 Qy  
907 GAGCTCAGCTCAGTCACCGTGGAGAAAGCCAGCTGCTCACAGTCACCCCGGGAGTACA 966  
301 GluHisSerIleProThrProThrSerAlaAlaProSerGluSerThrProSerGlu 320  
320 Qy  
967 GAGCAGCATCCCAACACCTCCCATAGCCAGCCCTCTGAGTCCACCCCATCTGAG 1026  
321 LeuProIleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAlaGly 340  
1027 CTACCCATATCTCTTACCACCTGCTCCCGAGCAGTGAAGAACTTACGGTATCGCTGA 1086  
341 AspAsnLeuIleIleThrLeuProAspAsnGluValGluLeuLysAlaPheValAlaPro 360  
360 Qy  
1087 GATAACCTTAATTATAACTTTTACCCGACAATGAAGTTGAACCTTGAAGGCCCTTTGTGCGCCA 1146  
361 AlaProProValGluThrThrTyrAsnTyrGluTyrAsnLeuIleSerHisProThrAsp 380  
380 Qy  
1147 GCGCCACCTGTAGAAACCACTTACAACTATGAATGAATTTAATAAGCCACCCACAGAC 1206  
381 TyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerVal 400  
400 Qy  
1207 TACCAAGGTGAATATAAACAAGGACACCAAGCAACTCTTAACCTCTCTCAATTTGTCGTC 1266  
401 GlyLeuTyrValPheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheVal 420  
420 Qy  
1267 GGACTTTATGCTTCAAAAGTCACTGTTTCTAGTGAAAACGGCTTTGGAGAGGATTTGTC 1326  
421 AsnValThrValLysProAlaArgArgValAsnLeuProProValAlaValValSerPro 440  
440 Qy  
1327 AATGTCACTGTAAAGCCCTGCCAGAGAGTCAACCTGCCACCTGTAGCAGTGTGTTCTCCC 1386  
441 GlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThr 460  
460 Qy  
1387 CACTGCAAGAGCTCACTTTGCTTTGACGTGACCCCTCATTTGATGGCAGCCCAAGTACA 1446  
461 AspAspThrGluIleValSerTyrHisTrpGluGluIleAsnGlyProPheIleGluGlu 480

1447 GATGATACTGAATAGTAGTATCATTTGGGAAGAAATAAACGGGCCCTTCTATAGAGAG 1506  
481 LysThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSer 500  
500 Qy  
1507 AAGACTTCAGTTGACTCTCCGCTTCTAACCCTTGATCTCTGGTAACTATAGT 1566  
501 PheArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaLeuLeu 520  
520 Qy  
1567 TTCAGGTGACTGTTCAGACTCGGAGGAGCCACTAACTCTACAACATGACAGCCCTTAATA 1626  
521 ValAsnAsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThr 540  
540 Qy  
1627 GTGAAACAATGCTGTGGACTACCCACAGTTGCTTAATGCAGGACCAATCACACATAACT 1686  
541 LeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleVal 560  
560 Qy  
1687 TTGCCCAAAACTCCATCACTTTGAATGGAAACAGAGCAGTGCAGTACACAGATTGTC 1746  
561 LeuTyrGluTyrSerLeuGlyProGlySerGluGlyLysHisValMetGlnGlyVal 580  
580 Qy  
1747 CTCTATCAGTGTCTCCCTGGGTCTGGGAGTGGGGCAACATGTGGTCTATGCAGGGAGTA 1806  
581 GlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLys 600  
600 Qy  
1807 CAGACGCATACCTTCACTTATCTGCAATGCAGAGGAGATATATACATTTTCAGCTGAAG 1866  
601 ValThrAspSerSerArgGlnSerThrAla\*\*\*ValThrValIleValGlnProGlu 620  
620 Qy  
1867 GTGACAGATTCTTCAAGGCAACAGTCTACTCTGTGGTGACTGTGATTGTCCAGCCTGAA 1926  
621 AsnAsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGlu 640  
640 Qy  
1927 AACAAATAGACTCCAGTGGCTGTGGCGCCCTCATATAAGAGCTGATCTTCCCACTGAAA 1986  
641 SerAlaThrLeuAspGlySerSerSerAspAspHisGlyIleValPheTyrHisTrp 660  
660 Qy  
1987 AGTCTACCTTGGATGGAGCAGCAGCAGCATGACCAAGCATTTGCTTTTACCACTGG 2046  
661 GluHisValArgGlyProSerAlaValGluMetGluAsnIleAspLysAlaIleAlaThr 680  
680 Qy  
2047 GAGCAGCTCAGAGCCCTCAGTGAGTGAGATGAAATAATTTGACAAAGCAATAGCCACT 2106  
681 ValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGly 700  
700 Qy  
2107 GTGACTGTCTCCAGTGGGACCTTACCACTTCCGTTTGACAGTGAAGACCCAGCAGGA 2166  
701 LeuSerSerThrSerThrLeuThrValAlaValLysLysGluAsnAsnSerProProArg 720  
720 Qy  
2167 CTGAGCAGCAGCTCCACCTCAGTGGCTGTGAAGAGGAAATAATAGTCTCTCCCAAGA 2226  
721 AlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGly 740  
740 Qy  
2227 GCGCGGCTGTGGCAGCAGCATGTTCTGTGCTTCCCAATAATTTCCATTACTTTGGATGTT 2286  
741 SerArgSerThrAspAspGlnArgIleValSerTyrLeuTyrIleArgAspGlyGlnSer 760  
760 Qy  
2287 TCAAGGTCTACTGATGACCAAAAGAAATTTGTGCTTCTGTGGATCCGGATGGCCAGAGT 2346  
761 ProAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsn 780  
780 Qy  
2347 CCACAGCTGGAGATGTCATCGATGGCTCTGACCACAGTGTGGCTCTCGAGCTTACGAAT 2406  
781 LeuValGluGlyValTyrThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAsp 800  
800 Qy  
2407 CTGTTGGAGGGGTGTACACTTTCCATCTTGGAGTCCACGACAGTCCAGGGGGCTCGAC 2466  
801 ThrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValGluLeu 820  
820 Qy  
2467 ACAGACACTGCCACTGTGGAAGTGCAGCCAGCAGCCCTAGGAAGAGTGGCTCTGGTGGAGCTG 2526  
821 ThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAspThrLeuValArgGln 840

Db 2527 ACCCTGCAGTTGGTGTGGCAGCTGACAGACGCGAAGGACACACCCCTGTGTGGGCGAG 2586  
 Qy 841 LeuAlaValLeuLeuAsnValLeuAspSerAspIleLysValGlnLysIleArgAlaHis 860  
 Db 2587 CTGGCTGTGCTGCTGAACGTGCTGACCTCGGACATTAAGGTCCAGAAATTCGGGCCAC 2646  
 Qy 861 SerAspLeuSerThrValIleValPheTyrValGlnSerArgProPheLysValLeu 880  
 Db 2647 TCGGATCTCAGCACCGTGATCTGTTTATGTATGATGACAGACGCGCCCTTTCAGGTCTC 2706  
 Qy 881 LysAlaIaGluValAlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAspPhe 900  
 Db 2707 AAAGCTGTGAGTGGCCGGAATCTGCACATGCGGCTCTCAAGGAGAAAGCTGACTTC 2766  
 Qy 901 LeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeuLysCysSerGlyHis 920  
 Db 2767 TTGCTTTTCAAGGTCTTGAAGGTTCATACAGCAGGTTCCTTCTGAAGTGTTCGGCCAT 2826  
 Qy 921 GlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeuTrpMetGluAsnLeu 940  
 Db 2827 GGTCACTGCGACCCCTCACAAGCGCTGCATTTGCTCTCACTTATGATGGAGAACCTT 2886  
 Qy 941 IleGlnArgTyrIleTrpAspGlyGluSerAsnCysGluTrpSerIlePheTyrValThr 960  
 Db 2887 ATACAGCGTTATATCTGGGATGGAGAGCAACTGTGA----- 2924  
 Qy 961 ValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrpLeuCysIleCysCys 980  
 Db 2924 ----- 2924  
 Qy 981 CysLysArgGlnLysArgThrLysIleArgLysLysThrLysTyrThrIleLeuAspAsn 1000  
 Db 2925 -----CAAAAAGGACTAAATCAGAAAAAACAAGTACACCACTCTGGATAAC 2975  
 Qy 1001 MetAspGluGlnGluArgMetGluLeuArgProLysTyrGlyIleLysHisArgSerThr 1020  
 Db 2976 ATGGATGAACAGAAAGATGNAACTGAGGCCCAATATGTTATCAAGCAGCAGGAGCACA 3035  
 Qy 1021 GluHisAsnSerSerLeuMetValSerGluSerGluPheAspSerAspGlnAspThrIle 1040  
 Db 3036 GAGCAAACTCCAGCGCTGATGGTATCCGAGTCTGAGTTTGACAGTGACAGGACCAATC 3095  
 Qy 1041 PheSerArgGluLysMetGluArgGlyAsnProLysValSerMetAsnGlySerIleArg 1060  
 Db 3096 TTCAGCCGAGAAAGATGAGAGAGGGGATCCAAAGGTTTCCATGATGTTCCATCAGA 3155  
 Qy 1061 AsnGlyAlaSerPheSerTyrCysSerLysAspArg 1072  
 Db 3156 AATGGAGCTTCCTTCAGTTATTGCTCAAGGACAGA 3191  
 RESULT 8  
 ADM36226  
 ID ADM36226 standard; cDNA; 3147 BP.  
 AC ADM36226;  
 XX  
 DT 03-JUN-2004 (first entry)  
 DE Murine SST20-6 cDNA related to adipocyte function SeqID 21.  
 KW gene; ss; mouse; murine; differentiation; metabolic function; adipocyte;  
 KW obesity; hypertension; hyperlipaemia; diabetes; arteriosclerosis;  
 KW anorectic; antidiabetic; antiarteriosclerotic; antilipaemic; hypotensive.  
 XX Mus musculus.  
 OS  
 XX  
 XX WO2004007711-A1.  
 XX  
 XX 22-JAN-2004.  
 PD  
 XX  
 PF 09-JUL-2003; 2003WO-JP008690.  
 XX  
 PR 10-JUL-2002; 2002JP-00201856.

XX (TAKE ) TAKEDA CHEM IND LTD.  
 PA Matsuzawa Y, Funahashi T, Shimomura C, Furuyama N;  
 PI WPI; 2004-122943/12.  
 DR P-PSDB; ADM36227.  
 XX Mouse membrane and secretory proteins of adipocyte origin and  
 PT polynucleotides encoding them for screening compounds as remedies for  
 PT obesity, diabetes, arteriosclerosis, hypertension and hyperlipemia.  
 XX  
 PS Disclosure; SEQ ID NO 21; 195pp; Japanese.  
 XX  
 CC This invention relates to novel membrane proteins that are associated  
 CC with differentiation and/ or metabolic function of adipocytes, in  
 CC particular of mouse origin. Specifically, it refers to the isolated  
 CC nucleic acid molecules encoding all or part of these proteins,  
 CC appropriate antibodies and screening assays useful for the development of  
 CC drug compositions derived thereof. The present invention describes these  
 CC compositions as useful for the treatment of diseases associated with  
 CC abnormalities of adipocyte function, such that they can be used to  
 CC prevent, treat or diagnose obesity, hypertension, hyperlipaemia, diabetes  
 CC and arteriosclerosis. Accordingly, they exhibit anorectic, antidiabetic,  
 CC antiarteriosclerotic, antilipaemic and hypotensive activities. This  
 CC polynucleotide is a murine cDNA sequence encoding a membrane protein of  
 CC the invention.  
 XX  
 SQ Sequence 3147 BP; 816 A; 867 C; 843 G; 621 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 4.85e-138 Length: 3147  
 Score: 2435.50 Matches: 528  
 Percent Similarity: 61.87% Conservative: 139  
 Best local Similarity: 48.98% Mismatches: 303  
 Query Match: 43.65% Indels: 109  
 DB: 12 Gaps: 19  
 US-10-764-390-3 (1-1072) x ADM36226 (1-3147)  
 Qy 10 SerLeuLeuLeuValThrIleAlaGlyCysAlaArgLysGlnCysSerGluGlyArg 29  
 Db 121 AGCGTCTGTGGTGTTCACAGATGCTGATGAGACAGA-----TGCCACAGGGGAAG 174  
 Qy 30 ThrTyrSerAsnAlaValIleSerProAsnLeuGluThrThrArgIleMetArgValSer 49  
 Db 175 ACACCTTATGAGCTGGCTTGAGAACTGAGGAGAA---AATCAGCTCGGGTCTTGTGA 231  
 Qy 50 HisthrPheProValValAspCysThrAlaAlaCysCysAspLeuSerCysAspLeu 69  
 Db 232 GGAAGCCTGCCTTTCACAGCCTGTGCGGCTGCTGCTGCGGAGACTTGCCTGCCACGCT 291  
 Qy 70 AlaTrpTrpPheGluGlyArgCysTyrLeuValSerCysProHisLysGluAsnCysGlu 89  
 Db 292 CTATGCTGGCTGGAAGGATGCTTTCAGGCTGACTGACGCAAGCCCGCAGAGTCCAG 351  
 Qy 90 ProLysLysMetGlyProIleArgSerTyrLeuThrPheValLeuArgProValGlnArg 109  
 Db 352 CTTTTAGACAGACTCTTCCATTCCATGCTG----- 384  
 Qy 110 ProAlaGlnLeuLeuAspTyrGlyAspMetMetLeuAsnArgLysSerProSerGlyIle 129  
 Db 385 -----ATCATTTTCAAAAAATCCCAAACTACAGATGAT 417  
 Qy 130 TrpGlyAspSerProGluAspIleArgLysAspLeu\*\*\*PheLeuGlyLysAspTrpGly 149  
 Db 418 TTGGGCTTCTGCTGGAAGATGATGAACACATCTTCTGAGGTAGGC-----TGGGGC 471  
 Qy 150 LeuGluGluMetSerGluTyr\*\*\*AspAspTyrArgGluLeuGluLysAspLeuGln 169  
 Db 472 AG-GACATCGTGGAGAG-----GCAGAG 494  
 Qy 170 ProSer---GlyLysGlnGluPro-----Arg 177



QY 887 gAenLeuHisMetArgLeuSerLysGluLysAlaAspPheLeuLeuPheLysValLeuAr 907  
 Db 2580 CATGCTCAAGCAGAGCTGCAGACAGAGAGCTACTCTCTCATCTTCAGAGCCCTGGA 2639  
 QY 907 gValAspThrAlaGlyCysLeuLeuLysCysSerGlyHisGlyHisCysAspProLeuTh 927  
 Db 2640 AATCAGCACAGTCACATGTGAGCTGAAGCTGTTCTGACCATGCCACTGTGACTCATTCAC 2699  
 QY 927 rlyAspGlyCysLysCysSerHisLeuTrpMetGluAsnLeuLeuGlnArgTyrIleTrpAs 947  
 Db 2700 CAAGCGCTGTGCTGTGAGCCCTTTTGGATGGAGAAATTCATCAAGGTGCAGCTGAGGGA 2759  
 QY 947 pGlyGluSerAenCysGluTrpSerIlePheTyrValThrValLeuAlaPheThrLeuIl 967  
 Db 2760 TGGACAGACAACTGTGAATGAGAGGTGCTTACGTATCATCTGCTCTCTTGTCTATGT 2819  
 QY 967 eValLeuThrGlyGlyPheThrTrpLeuCysIleCysCysLysArgGlnLysArgTh 987  
 Db 2820 TGTTCCTTGGGGATCCTGTATGACTACAACTGCTGCTGCAAGAGGCAAAAA--GG 2876  
 QY 987 rlyLysLeuArgLysLysThrLysTyrThrIleLeuAspAsnMetAspGluGlnLysArgMe 1007  
 Db 2877 AAAACCCCAAGAGAAAGCAGATACAAAGATCCTGGATGCCAGAT--CAGGAGAGCCT 2933  
 QY 1007 tGluLeuArgPro-----LysTyrGlyIleLysHisArgSerThrGluHisAsnSerSe 1025  
 Db 2934 GGAGCTGAACCAACCTCCGAGAGGAGGAGCAACAGAAAGGCCCCACGCTGACGAGCAG 2993  
 QY 1025 rLeuMetValSerGluSerGluPheAspSerAspGlnAspThrIlePheSerArgGluLyl 1045  
 Db 2994 CCTGATGCATCTGATCGAGCTGGACAGCGAC--GATGCCATCTTCACATGCCGAGA 3050  
 QY 1045 eMetGluArgGlyAsnProLysValSerMetAsnGlySerIleArgAsnGly 1062  
 Db 3051 CGGGAGAGAGGGCAAACTACTGTATGTCAGAAATGCTCTGTGCCAAAGCGG 3102  
 RESULT 9  
 ADI21796  
 ID ADI21796 standard; cDNA; 3846 BP.  
 AC  
 XX ADI21796;  
 DT  
 XX 15-APR-2004 (first entry)  
 DE  
 XX Novel human protein cDNA #55.  
 KW forensic; nutritional source; damaged tissue; diseased tissue;  
 KW myeloid cell disorder; lymphoid cell disorder;  
 KW bone cartilage tissue growth; tendon tissue growth;  
 KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;  
 KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human;  
 ss; Gene.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003025148-A2.  
 XX  
 PD 27-MAR-2003.  
 XX  
 XX 19-SEP-2002; 2002WO-US029964.  
 PF  
 XX 19-SEP-2001; 2001US-0323739P.  
 PR 13-SEP-2002; 2002US-00323739.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;  
 PI Haley-Vicente D;  
 XX WPI; 2003-354603/33.  
 DR P-PSDB; ADI21080.

XX New polynucleotides and secreted proteins, useful for treating myeloid or  
 PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
 PT tissue growth or regeneration, in wound healing, and in tissue repair and  
 PT replacement.  
 XX Claim 1; SEQ ID NO 55; 156pp; English.  
 XX The invention relates to an isolated polynucleotide encoding a  
 CC polypeptide with biological activity. The polynucleotides and  
 CC polypeptides are useful in diagnostics, forensics, gene mapping,  
 CC identification of mutations responsible for genetic disorders and other  
 CC traits, to assess biodiversity, as nutritional sources or supplements.  
 CC The polynucleotides may also be used as molecular weight markers,  
 CC chromosome markers or map related gene positions, or as an antigen to  
 CC raise anti-DNA antibodies or elicit immune response. The polypeptides are  
 CC useful for raising antibodies, as markers for tissues in which the  
 CC corresponding polypeptide is expressed, for re-engineering damaged or  
 CC diseased tissues, for treating myeloid or lymphoid cell disorders, in  
 CC bone cartilage, tendon, ligament and/or nerve tissue growth or  
 CC regeneration, in wound healing, in tissue repair and replacement, in  
 CC healing of burns, incisions and ulcers, and in treating cancer. The  
 CC present sequence represents cDNA encoding a novel human protein.  
 XX  
 SQ Sequence 3846 BP; 1008 A; 1005 C; 991 G; 842 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 5.02e-136 Length: 3846  
 Score: 2404.00 Matches: 522  
 Percent Similarity: 60.13% Conservative: 134  
 Best Local Similarity: 47.85% Mismatches: 318  
 Query Match: 43.08% Indels: 117  
 DB: 10 Gaps: 17  
 US-10-764-390-3 (1-1072) x ADI21796 (1-3846)  
 QY 4 ProThrGlyValLeuSerSerLeuLeu-----LeuValThrIleAlaGlyCys-Al 21  
 Db 280 CCTGTACTCTGTTTATCTGCTTTTGTCTTCAGCGTCTGTGTTGTCAACAGATGCCAG 339  
 QY 21 aArgLysGlnCysSerGluGlyArgThrTyrSerAsnAlaValIleSerProAsnLeuGl 41  
 Db 340 TCAGAGCAGGTGCCAGCAGGGGAAAGACACAATTTGGAGTTGGCCTCGAGATCTGGGGGAGA 399  
 QY 41 uThrThrArgIleMetArgValSerHisThrPheProValValAspCysThrAlaAlaCy 61  
 Db 400 A---AATCACCTCTGGCTTCTTGAAGGAACCCCTCTCTCAGTCATGTTGGCTGCCTG 456  
 QY 61 sCysAspLeuSerSerCysAspLeuAlaTrpPheGluGlyArgCysTyrLeuValse 81  
 Db 457 CTGCCAGGACTCTGCCTGCCATGCTTTTGTGGCTAGAGGATGTGCATTCAGGCAGA 516  
 QY 81 rCysProHisLysGluAsnCysGluProLysLysMetGlyProIleArgSerTyrLeuTh 101  
 Db 517 CTGCAGCAGGCCCCAGAGCTGCCGGGCTTTTAGGACACACTCCTCCAATTCATGCTGCT 576  
 QY 101 rPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetMetle 121  
 Db 577 GTTT---TTAAAAAAATTCCAAACTGCAGATGATTG----- 610  
 QY 121 uAsnArgGlySerProSerGlyIleTrpGlyAspSerProGluAspIleArgLysAspLe 141  
 Db 611 -----GGCTTTTCTACCTGAAGATGATGTACCATCTCT 642  
 QY 141 u\*\*\*PheLeuGlyLysAspTrpGlyLeuGluGluMetSerGluTyr\*\*\*AspAspTyrAr 161  
 Db 643 TCTGGGGCTAGGTGGAACTGGGCA----- 667  
 QY 161 gGluLeuGluLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAlaGl 181  
 Db 668 -----TCTTGGAGGCGAGAGCCCAACCCAGAGCTGCAC 699  
 QY 181 uTyrThrAspTrpGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGlyAs 201



Db 700 C----- 700  
Qy 201 pSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGluSe 221  
Db 701 -AGACCTGCTGTATCTCCAGTCAGCCAGCAG----- 730  
Qy 221 rAlaSerThrProAlaProLysLeuProGluArg-----SerValLeuLeuPr 237  
Db 731 -----AGCTTAATCAGAAAGAGAGGTAGTCCCGAGTCAGCTAGTTACACC 783  
Qy 237 oLeuProThrThrProSerSerGlyGluValLeuGluLysAlaSerGlnLeuGl 257  
Db 784 TATAGTGACAGCATCT-----AAAGTGAATGACTCCCAAGAAATTAGTGTCTGAC 837  
Qy 257 nGluGlnSerSerAsnSerSerGlyLysGluValLeuMetProSerHisSerLeuProPr 277  
Db 838 TACCAGTGGCTCTGCAGAGGTCACCAAGCGATTACAAATTTCCAGTCCCTTAACCACAGA 897  
Qy 277 oAlaSerLeuGluLeuSer-----SerValThrValGluLysSerPr 291  
Db 898 CTGACTGCAGAGCTGTCTGTGGGCCAAAGAATGTATCAGTGCACCAACCTGAAATATCAGA 957  
Qy 291 oValLeuThrValThrProGlySer-----ThrGluHisSe 303  
Db 958 GGGTCTTGCTACTACGCCAGCAGCTCAACAAGTAAAGTTCTGAGAAAACCCAGATTGC 1017  
Qy 303 rIleProThrPro-----ProThrSerAlaIaIaProSe 314  
Db 1018 TGTCCCCAGCCAGTGGCTCCTCTACAGTTATGTACTACCCCTACCCCGCCAGGCTCTTT 1077  
Qy 314 rGluSerThrProSerGluLeuProIleSerProThrThrAlaProArgThrValLysGl 334  
Db 1078 CCAGAGCACCTCAGCACCATACCCAGTT-----ATAAAGGA 1113  
Qy 334 uLeuThrValSerAlaGlyAspAsnLeuIleIleThrLeuProAspAsnGluValGluLe 354  
Db 1114 ACTGGTGTATCTGTGGAGAGGTCCAGATACCCCTGCCTAAGAAATGAATTCATTT 1173  
Qy 354 uLysAlaPheValAlaProAlaProValGluThrThrTyrAsnTyrGluThrProAsnLe 374  
Db 1174 AAATGCATATGTTCTCCAAGAACCCACTAAAGGAGAAACCTACACCTACGACTGGCAGCT 1233  
Qy 374 uIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAs 394  
Db 1234 GATTACTCATCTAGACACTACGTGGAGAAATGGAAGGAAACATTTCCAGATCTCTCAA 1293  
Qy 394 nLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSerGluAsnAl 414  
Db 1294 ACTATCGAAGCTCACTCCAGGCCTGTATGAATTCAAAGTGATTTAGAGGGTCAAAATGC 1353  
Qy 414 aPheGlyGluGlyPheValAsnValThrValLysProAlaArgValAsnLeuProPr 434  
Db 1354 CCATGGGGAAGGCTATGTGAAGTGCAGCTCAGCCAGAGCCCGGTAAGAAATCGGCCCC 1413  
Qy 434 oValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIl 454  
Db 1414 CATTGCTATTGTGCACCTCAGTTCAGGAGATCTCTTTGGCAACCACTTTCTACAGTCAT 1473  
Qy 454 eAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisTyrGluGluLeuAs 474  
Db 1474 TGATGGCAGTCAAAAGCACTGATGATGAATAATCGTTCCAGTACCATTGGGAAGAACTTAA 1533  
Qy 474 nGlyProPheIleGluGluLysThrSerValAspSerProValLeuArgLeuSerAsnLe 494  
Db 1534 GGGGCTCTAAGAGAGAGAGAGATTTCTGAAGATACAGCCATATTAACCTAAGTAAACT 1593  
Qy 494 uAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSe 514  
Db 1594 GGTCCCTGGGAACCTACACTTTTCAGCTGTAGTACTCTGATGAGCTACCAACTC 1653  
Qy 514 rThrThrAlaIaLeuIleValAsnAsnAlaValAspTyrProProValAlaAsnAlaGl 534

Db 1654 TACTACTGCAAACTGACAGTGAACAAAGCTGTGGATTACCCCTCTGGCCCAAGCAGG 1713  
Qy 534 yProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSe 554  
Db 1714 CCCCAACCAAGTATGATCACTCCCTGCCCCAAAACTCCATCACCTCTTTTGGGAACCAAGCAGCAC 1773  
Qy 554 rAspAspHisGlnIleValLeuTyrGluTyrSerLeuGlyProGlySerGluGlyLysHi 574  
Db 1774 TGATGATCATGGCATCACAGCTATGATGTGTCTACTCAGCCCAAGCAGCAAGGGAAGT 1833  
Qy 574 sValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAs 594  
Db 1834 GGTGAGATGCAGGGTGTAGAACACCAACCTTACAGCTCTCTCGGATGCAAGAAGGAGA 1893  
Qy 594 pTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValTh 614  
Db 1894 CTACACTTACAGCTCAGAGTACTGACACAAATAGGACAGCAGGCGCACTGTCTCAAGTGAC 1953  
Qy 614 rValIleValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLysGl 634  
Db 1954 TGTATTGTGCAACCTGAAACCAATAAGCCTCTCTCAGCGAGATCGAGCCCGAGATAAAGA 2013  
Qy 634 uLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerSerAspAspHisGl 654  
Db 2014 GCTGACCTTCTGTGGATAGCACCAACCTGGATGGCAGCAAGAGCTCAGATGATCAGAA 2073  
Qy 654 yIleValPheTyrHisTyrGluHisValArgGlyProSerAlaValGluMetGluAsnIl 674  
Db 2074 AATTATCTCATCTCTGGGAAAAAACACAGGGACCTGATGGGGTGCAGCTCGAGAAATGC 2133  
Qy 674 eAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuTh 694  
Db 2134 TAACAGCAGTGTCTACTGTGCTGGCTGCAAGTGGGGACCTATGTGTTCACTTGAC 2193  
Qy 694 rValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysLysGl 714  
Db 2194 TGTCAAGATGAGAGGAACTGCAAGCCAGAGCTCTGTGAATCTCATTTGTCAAGAAGA 2253  
Qy 714 uAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAs 734  
Db 2254 AATAAAACAACCACTATAGCAAGATAACTGGGAATGTGTGTGATTACCTTACCCACAG 2313  
Qy 734 nSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuTr 754  
Db 2314 CACAGCAGCTGGATGGCTCTAAGTCTCTCAGATGACAAGGGAATAGTCACTCACTCTG 2373  
Qy 754 pIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerVa 774  
Db 2374 GACTCGAGATGAGGGGAGCCAGCAGCAGGGAGGTGTAAATCACTCTGACCATCACCC 2433  
Qy 774 lAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrAs 794  
Db 2434 TATCTTTTCTTTCAAACTGGTTGAGGGAACCTACATTTTCCCTGAAAGTGACCGA 2493  
Qy 794 pSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArgLys 814  
Db 2494 TGCAAGGGTGAGAGTGACACAGCCGACCACTGTGGAGGTGAACCTGTATCCAGGAA 2553  
Qy 814 sSerGlyValGluLeuLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLys 834  
Db 2554 AAACAACCTGGTGAGATCATCTTGGATATACAGTCAGTCAGCTAAGTCTGAGAGGCTGAA 2613  
Qy 834 sAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLysVa 854  
Db 2614 GGGGATGTTCACTCCCGCAGATGGGGTCTCTCTGGGGTGTGTGATTTCCGACATCATTTGT 2673  
Qy 854 lGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSerAr 874  
Db 2674 GCAAAAGATTCAGCCGTCACAGGAGCAGCAGCAAAATGTGTATTTTGTTCAAAACGA 2733  
Qy 874 gProProPheLysValLeuLysAlaAlaGluAlaIaArgAsnLeuHisMetArgLeuSe 894  
Db 2734 GCCTCCCCACCAATCTTCAAAAGGCCCATGAGGTGGCAGCGATGTCTCAAGAGTGAGCTCG 2793



Db 533 CTGCAGAGGCCCCAGAGCTGCCGGGCTTTAGACACACACTCTCCCAATTCCATGCTGGT 592  
Qy 101 rPheValLeuArgProValGlnArgProAlaGlnLeuLeuAspTyrGlyAspMetLe 121  
Db 593 GTTTT---TTAAAAAATTCAAACTGCAGATGATTG----- 626  
Qy 121 uAsnArgGlySerProSerGlyIleTyrGlyAspSerProGluAspIleArgLysAspLe 141  
Db 627 -----GGCTTTCTACCTGAAGATGATGATACCATCT 658  
Qy 141 u\*\*\*PheLeuGlyLysAspTyrGlyLeuGluGluMetSerGluTyr\*\*\*AspAspTyrAr 161  
Db 659 TCTGGGCTAGTTGGAACTGGGA----- 683  
Qy 161 gGluLeuGlyLysAspLeuLeuGlnProSerGlyLysGlnGluProArgGlySerAlaG 181  
Db 684 -----TCTTGAGGAGCAGAGCCACCCAGAGCTGCAC 715  
Qy 181 uTyrThrAspTyrGlyLeuLeuProGlySerGluGlyAlaPheAsnSerSerValGlyAs 201  
Db 716 C----- 716  
Qy 201 pSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAsnGluSe 221  
Db 717 -AGACTTGCTGTATCTCCAGTGACACGAG----- 746  
Qy 221 rAlaSerThrProAlaProLysLeuProGluArg-----SerValLeuLeuPr 237  
Db 747 -----AGCTTAATCAGGAAGCTTCAGAGAGAGGTAGTCCCAAGTGCAGTAGTTACACC 799  
Qy 237 oLeuProThrThrProSerSerGlyGluValLeuGluLysGlyLysAlaSerGlnLeuG 257  
Db 800 TATAGTCACACAGCATTTCT-----AAAGTGAATGACTCCCAACGAATTAGTGCTGCAC 853  
Qy 257 nGluGlnSerSerAsnSerGlyLysGluValLeuMetProSerHisSerLeuProPr 277  
Db 854 TACAGAGTGGCTCTGCAGAGTCCACAGGCGATACAAATTTCCAGTCCCTTAACACACAGA 913  
Qy 277 oAlaSerLeuLeuLeuSer-----SerValThrValGluLysSerPr 291  
Db 914 CTTGACTGCAGAGCTGTCTGTGGGCGCCAAAGATGATCAGTGCACCTGAATATACAGA 973  
Qy 291 oValLeuThrValThrProGlySer-----ThrGluHisSe 303  
Db 974 GGGTCTTGCTACTAGCGCCAGCACTCAACAGTAAAAAAGTTCTCAGAAAAACCCAGATTGC 1033  
Qy 303 rIleProThrPro-----ProThrSerAlaAlaProSe 314  
Db 1034 TGTCCCCCAGCCAGTGGCTCCTCTCAGTTATGCTACCCCTACCCCGCCAGGCTCTTT 1093  
Qy 314 rGluSerThrProSerGluLeuProIleSerProThrThrAlaProArgThrValLysG 334  
Db 1094 CCAGAGCACCTCAGCACCATACCCAGTT-----ATAAGGA 1129  
Qy 334 uLeuThrValSerAlaGlyAspAsnLeuIleThrLeuProAspAsnGluValGluLe 354  
Db 1130 ACTGGTGTAATCTCTGGAGAGAGTGTCCAGATAACCTCGCTTAAGAAATGAATTCAT 1189  
Qy 354 uLysAlaPheValAlaProAlaProProValGluThrThrTyrAsnTyrGluTyrAsnLe 374  
Db 1190 AAATGCAATATGTTCTCCAAGAACCCACTAAAGGAGAAACCTTACACCTACGACTGGCAGCT 1249  
Qy 374 uIleSerHisProThrAspTyrGlnGlyGluIleLysGlnGlyHisLysGlnThrLeuAs 394  
Db 1250 GATTACTCATCTAGAGACTACAGTGGAGAAATGGAAGGAAACATTCCTCCAGATCTCTCA 1309  
Qy 394 nLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSerGluAsnAl 414  
Db 1310 ACTATCGAAGCTCACTCCAGGCTGTATGAATTCAAAGTGTAGAGGGTCAAAATGC 1369  
Qy 414 aPheGlyGluCysPheValAsnValThrValLysProAlaArgValAsnLeuProPr 434

Db 1370 CCATGGGGAAGGCTATGTGAACGTGACAGTCAAGCCAGAGCCCGTGAAGATCGGCCCC 1429  
Qy 434 oValAlaValValSerProGlnLeuGlnLeuLeuThrLeuProLeuThrSerAlaLeuI 454  
Db 1430 CATTTGCTATTGTGTACCTCAGTTCCAGAGAGATCTCTTTGCCAACCACTCTTACAGTCAT 1489  
Qy 454 eAspGlySerGlnSerThrAspAspThrGluIleValSerTyrHisThrGluGluLeAs 474  
Db 1490 TGATGGCAGTCAAAAGCACTGATGATGATAAATCGTTAGTACCATTTGGGAGAACTTAA 1549  
Qy 474 nGlyProPheIleGluGluLysThrSerValAspSerProValLeuArgLeuSerAsnLe 494  
Db 1550 GGGGCTCTAAGAGAGAGAGATTCTGAAGATACAGCCATATTAACACTAAGTAACT 1609  
Qy 494 uAspProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAsnSe 514  
Db 1610 CGTCCCTGGGAACCTACACTTTACCTGTAGTAGACTCTGTATGAGACTCTGATGGAGTACCAATC 1669  
Qy 514 rThrThrAlaAlaLeuIleValAsnAsnAlaValAspTyrProProValAlaAlaAlaG 534  
Db 1670 TACTACTGCAAACTGACAGTGAACAAAGCTGTGGATTACCCCTCTGGCCAAAGCAGG 1729  
Qy 534 YProAsnHisThrIleThrLeuProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSe 554  
Db 1730 CCCCAACCAAGTGATCACCCTGCCCAAACTCCATCACCTCTTTGGGAACCAAGAGCAC 1789  
Qy 554 rAspAspHisGlnIleValLeuTyrGluTyrSerLeuGlyProGlySerGluGlyLysHi 574  
Db 1790 TGATGATCATGGCATCACAGCTATGATGTGTCTCACTCAGCCCAAGCAGCAAGGGAAGT 1849  
Qy 574 sValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAs 594  
Db 1850 GGTGAGATGAGGGTGTAGAACCAACCTTACAGCTCTCTCGCATGCAAGGAAGAGA 1909  
Qy 594 pTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValTh 614  
Db 1910 CTACACTTACCACTGCTACAGTGCACATAGGACAGCAGCCACTGCTCAAGTGAC 1969  
Qy 614 rValIleValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLysG 634  
Db 1970 TGTATTGTGCACCTGMAAACATAAGCTCTCTCAGCAGATGTCAGGCCCCAGATAAGA 2029  
Qy 634 uLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerSerSerAspHisG 654  
Db 2030 GCTGACCTTCTCTGTGATAGCACACCTGGATGGCAGCAAGAGCTCAGATGATCAGAA 2089  
Qy 654 YIleValPheTyrHisThrGluHisValArgGlyProSerAlaValAlaMetGluAsnI 674  
Db 2090 AATTATCTCATNTCTTGGGAAAAACACAGGAGCTGTATGGGTGAGCTCGAGATGC 2149  
Qy 674 eAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuTh 694  
Db 2150 TAACAGCAGTGTGTCTACTGTACTGGGCTGCAAGTGGGACCTATGTGTTTCACTTGAC 2209  
Qy 694 rValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysG 714  
Db 2210 TGTCAAGATGAGAGGAACCTGCAAGCCAGAGCTCTGTGAATGTCTATTGTCAAAGAAGA 2269  
Qy 714 uAsnAsnSerProProArgAlaArgAlaGlyGlyArgHisValLeuValLeuProAsnAs 734  
Db 2270 AATAAACAAACCACTATAGCAAGATAACTGGGAATGTGTGATTACCTTACCACAGAG 2329  
Qy 734 nSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuTr 754  
Db 2330 CACAGCAGAGTGGTGGCTCTAAGTCTCAGATGACAGGGGAATAGTCAGTACCTCTG 2389  
Qy 754 pIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerVa 774  
Db 2390 GACTCGAGATGAGGGGAGCCAGCAGAGGGGAGGTGTAAATCACTCTGACCATCACCC 2449  
Qy 774 lAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrAs 794  
Db 2450 TATCTCTTTTCTTCAAACTGTGTGAGGGAACCTACACTTTTCCCTTGAAGTGACCGA 2509



Db 717 C----- 717  
Qy 201 pSerProAlaValProAlaGluThrGlnGlnAspProGluLeuHisTyrLeuAenGluSe 221  
Db 718 -AGACCTGCTGTATCTCCAGTGACGACGAG----- 747  
Qy 221 rAlaSerThrProAlaProLysLeuProGluArg-----SerValLeuLeuPr 237  
Db 748 -----AGCTTAATCAGAAAGCTTCAGAAAGAGAGGTAGTCCCGAGTGACGTAGTATACACC 800  
Qy 237 oLeuProThrThrProSerSerGlyGluValLeuGluLysGluLysAlaSerGlnLeuGl 257  
Db 801 TATAGTGACACAGTATCT-----AAAGTGAATGACTCCCAACGAATTAGTGCTCTGAC 854  
Qy 257 nGluGlnSerSerAenSerSerGlyLysGluValLeuMetProSerHisSerLeuProPr 277  
Db 855 TACCAGTGGCTCTGCAGAGGTCCACAGGCGAATTAACAATTTCCAGTCCCTTAACCCACAGA 914  
Qy 277 oAlaSerLeuGluLeuSer-----SerValThrValGluLysSerPr 291  
Db 915 CTTCAGCTGCAGAGCTGTCTGGTGGGCCCAAGAATGTATCAGTGCACACCTGAAATATCAGA 974  
Qy 291 oValLeuThrValThrProGlySer-----ThrGluHisSe 303  
Db 975 GGGTCTTGCTACTAGCCGACACATCAACAAGTAAAGATTCTGAGAAACCCAGATTGC 1034  
Qy 303 rIleProThrPro-----ProThrSerAlaAlaProSe 314  
Db 1035 TGTCCCGCAGCGAGTGGCTCCTCTCAGATTATGTCTACCCCTACCCCGGCGCTCTTT 1094  
Qy 314 rGluSerThrProSerGluLeuProLysSerProThrThrAlaProArgThrValLysGl 334  
Db 1095 CCAGAGCACCTTCAGCACCATACCCAGTT-----ATAAAGGA 1130  
Qy 334 uLeuThrValSerAlaGlyAspAenLeuIleIleThrLeuProAspAenGluValGluLe 354  
Db 1131 ACTGGTGTATCTCTGGAGAGAGTGTCCAGATTAACCTCTGCTAAGATGAATGAATTCATTT 1190  
Qy 354 uLysAlaPheValAlaProAlaProValGluThrThrTyrAenTyrGluThrAenLe 374  
Db 1191 AAATGCATATGTTCTCCAAGAACCCACCTAAAGGAGAAACCTACACCTACGACTGGCAGCT 1250  
Qy 374 uIleSerHisProThrAspTyrGlnGlyGluLysGlnGlyHisLysGlnThrLeuAs 394  
Db 1251 GATTACTCATCTAGACATACAGTGGAGAAATGGAAGGAAACATTCCACAGATCTCAA 1310  
Qy 394 nLeuSerGlnLeuSerValGlyLeuTyrValPheLysValThrValSerSerGluAenAl 414  
Db 1311 ACTATCGAAGCTCACTCCAGGCGCTGTATGAATTCAAAGTGATTGTAGAGGCTCAAAATGC 1370  
Qy 414 aPheGlyGluGlyPheValAenValThrValLysProAlaArgValAenLeuProPr 434  
Db 1371 CCATGGGGAGGCTATGTGAACGTGACAGTCAAGCCAGAGCCCGCTAAGATCGGCCCCC 1430  
Qy 434 oValAlaValValSerProGlnLeuGlnGluLeuThrLeuProLeuThrSerAlaLeuIl 454  
Db 1431 CATTTGCTATGTGCACCTCAGTTCAGGAGATCTCTTTGCGAACCACTTCTACAGTCAT 1490  
Qy 454 eAspGlySerGlnSerThrAspAspThrGluLysValSerTyrHisTTPGluGluLysAs 474  
Db 1491 TGATGGCAGTCAAAAGCACTGATGATGAATAATCGTTCAAGTACCATTTGGGAAGAACTTAA 1550  
Qy 474 nGlyProPheIleGluGluLysThrSerValAspSerProValLeuArgLeuSerAenLe 494  
Db 1551 GGGGCGCTCTAAGAGAGAGAGAGATTTCTGAAGATACAGCCATATTAACATGAAGTAACT 1610  
Qy 494 uAspProGlyAenTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAenSe 514  
Db 1611 CGTCCCTGGGAACATACACTTTTCAGCTTGACTGTAGTACTCTGATGAGCTACCAACTC 1670  
Qy 514 rThrThrAlaAlaLeuIleValAenAenAlaValAspTyrProProValAlaAenAlaGl 534

Db 1671 TACTACTGCAAACTGACAGTGAACAAAGCTGTGGATTACCCCTGTGGCCAAAGCAGG 1730  
Qy 534 yProAenHisThrIleThrLeuProGlnAenSerIleThrLeuAenGlyAenGlnSerSe 554  
Db 1731 CCCCAACCAAGTGTATCACCCTGCCCAAACTCCACCAACCCCTCTTTGGGAACCAAGAC 1790  
Qy 554 rAspAspHisGlnIleValLeuTyrGluTyrSerLeuGlyProGlySerGluGlyLysHi 574  
Db 1791 TGATGATCATGGCATCACAGCTATGAGTGTGCTACTCAGCCCAAGCAGCAAGGAAAGT 1850  
Qy 574 sValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAs 594  
Db 1851 GGTGAGATGTCAGGGTGTAGAACACCAACCTTACAGCTCTCTCGATGCAAGAGGAGA 1910  
Qy 594 pTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValTh 614  
Db 1911 CTACACTTACAGCTCAGAGTGTACACAATAGGACAGCAGGCGCACTGTCTCAAGTGAC 1970  
Qy 614 rValIleValGlnProGluAenAenArgProProValAlaValAlaGlyProAspLysGl 634  
Db 1971 CGTTATTGTGCAACCTGAAACCAATAAGCCCTCTCAGCGAGATGCGGCCCAGATAAAGA 2030  
Qy 634 uLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerSerAspHisGl 654  
Db 2031 GCTGACCTTCTGTGTAGTACCAACCTCGATGGCAGCAAGAGCTCAGATGATCAGAA 2090  
Qy 654 yIleValPheTyrHisTyrGlu-HisValArgGlyProSerAlaValGluMetGluAenI 674  
Db 2091 AATTATCTCATATCTCTGGGAAAAAACAACAGGACCTGATGGGTGCAGCTCGAAGATG 2150  
Qy 674 leAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuT 694  
Db 2151 CTAAACAGAGTGTGTACTGTGCTGGGTGCAAGTGGGACCTATGTGTTCACCTTGA 2210  
Qy 694 hrValLysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysLysG 714  
Db 2211 CTGTCAAGATGAGAGAACCTGCAAAGCCAGAGCTCTGTGAATGTCTATTGTCAAAGAAG 2270  
Qy 714 luAenAenSerProProAlaArgAlaGlyGlyArgHisValLeuValLeuProAenA 734  
Db 2271 AAATAAAACAAACCACTATAGCCAAGATACTGGGAATGTGGTATACCTACCCACCA 2330  
Qy 734 snSerIleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuT 754  
Db 2331 GCACAGCAGAGCTGGATGGCTCTAAGTCTCAGATGACAAGGGAATAGTACGTACCTCT 2390  
Qy 754 rPileArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerV 774  
Db 2391 GGACTCGAGATGAGGGAGGCCAGCAGAGGGAGGTGTAAATCACTCTGACCATCAACC 2450  
Qy 774 alAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThrA 794  
Db 2451 CTATCCTTTTCTTTCAAACTGTGTCAGGGAACCTACACTTTTCACTGAAAGTGACCG 2510  
Qy 794 spSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArgL 814  
Db 2511 ATGCAAAAGGTGAGAGTGACACAGACCGGACCACTGTGGAGGTGAAACCTGATCCCA 2570  
Qy 814 ysSerGlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgL 834  
Db 2571 AAAACAACTGGTGGAGATCATCTTTGGATATCAACAGTCAGTCAGCTAACTGAGAGCTGA 2630  
Qy 834 ysAspThrLeuValArgGlnLeuAlaValLeuLeuAenValLeuAspSerAspIleLysV 854  
Db 2631 AGGGGATGTTCAATCCCGCAGATTTGGGTCTCTCTGGGGGTCTGATTCGACATCATTG 2690  
Qy 854 alGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSerA 874  
Db 2691 TGCAAAAGATTTCAGCCCGTACCGGAGCAGACCAAAATGGTATTTTGTCTCAAAACG 2750  
Qy 874 rgProProPheLysValLeuLysAlaAlaGluValAlaArgAenLeuHisMetArgLeuS 894  
Db 2751 AGCTCTCCCAACAGATCTTCAAGGCCCATGAGGTGGCAGCGGATGCTCAAGAGTGAGCTGC 2810





Qy 190 GlySerGluGlyAlaPheAsnSerSerValGlyAspSerProAlaValProAlaGluThr 209  
Db 732 -----AGACCTGCTGTATCTTCCAGTGAC 755  
Qy 210 GlnGlnAspProGluLeuHisTyrLeuAsnGluSerAlaSerThrProAlaProLysLeu 229  
Db 756 CAGCAG-----AGCTTAATCAGGAGCTT 779  
Qy 230 ProGluArg-----SerValLeuLeuProLeuProThrThrProSerSerGly 245  
Db 780 CAGAGAGAGGTAGTCCAGTCAGTCAGTATACACTATAGTCACACAGCAATCT----- 833  
Qy 246 GluValLeuGluLysGluLysAlaSerGlnLeuGlnGlnSerSerAsnSerSerGly 265  
Db 834 AAGTGAATGACTCCAAAGAAATAGGTGGTCTGACTACAGTGGCTTCGACAGAGGTCCAC 893  
Qy 266 LysGluValLeuMetProSerHisSerLeuProProAlaSerLeuGluLeuSer----- 283  
Db 894 AAGCGATTACAATTTCCAGTCCCTTAACACAGACCTGACTGCAGAGCTGTCTGGTGG 953  
Qy 284 -----SerValThrValGluLysSerProValLeuThrValThrProGlySer 299  
Db 954 CCAAAGAATGTATCAGTCAACCTGAAATATCAGAGGGTCTTGCTACTACGCCCCAGCACT 1013  
Qy 300 -----ThrGluHisSerIleProThrPro----- 307  
Db 1014 CAACAAGTAAAGTTCTGAGAAACCCAGAAATGCTGTCCTCCAGCCAGTGGCTCCCTCC 1073  
Qy 308 -----ProThrSerAlaAlaProSerGluSerThrThrProSerGluLeuPro 322  
Db 1074 TACAGTTATGCTACCCCTACCCCGAGGCTCTTTCAGAGACCTCAGCACCATACCCA 1133  
Qy 323 IleSerProThrThrAlaProArgThrValLysGluLeuThrValSerAlaGlyAspAsn 342  
Db 1134 GTT-----ATAAGGAACCTGGTGTATCTGCTGGAGAGAGT 1169  
Qy 343 LeuIleIleThrLeuProAspAsnGluValGluLeuLysAlaPhe-ValAlaProAla-P 362  
Db 1170 GTCCAGATAACCCCTGCTTAAGAATGAAGTTCAATTAATGATATGTTCTTCAGGGAAC 1229  
Qy 362 roProValGluThrThrTyrAsnTyrGluThrProAsnLeuIleSerHisProThrAspTyrG 382  
Db 1230 CACCTAAAGGAGAAACCTACACCTACGACTGGAGCTGATCTATCTCTAGAGACTACA 1289  
Qy 382 InGlyGluIleLysGlnGlyHisLysGlnThrLeuAsnLeuSerGlnLeuSerValGlyL 402  
Db 1290 GTGGAGAAATGGAAGGGAACATTCCAGATCTCTCAAACTATCGAAGCTCACTCCAGGCC 1349  
Qy 402 euTyrValPheLysValThrValSerSerGluAsnAlaPheGlyGluGlyPheValAsnV 422  
Db 1350 TGTATGAATTCAAAGTGATTTAGAGGGTCAAAATGCCCATCGGGAAGGCTATGTGAACG 1409  
Qy 422 alThrValLysProAlaArgValAsnLeuProProValAlaValValSerProGlnL 442  
Db 1410 TGACAGTCAAGCCAGGCCCCGTAGAATCGGCCCTCCATTCGCTATTGTGCACCTCAGT 1469  
Qy 442 euGlnGluLeuThrLeuProLeuThrSerAlaLeuIleAspGlySerGlnSerThrAspA 462  
Db 1470 TCCAGGAGATCTCTTTGCCAACCACTTCTACAGTCAATTCATGGCAGTCAAAAGCACTGATG 1529  
Qy 462 spThrGluIle-ValSer-TyrHisTrrpGluGluIleAsnGlyProPheIleGluLys 481  
Db 1530 ATGATAAAATCCGTTTCAGTACCATTTGGGAAGAACTTAAAGGGGCCCTCTTAAGAGAAGAGAG 1589  
Qy 482 ThrSerValAspSerProValLeuArgLeuSerAsnLeuAspProGlyAsnTyrSerPhe 501  
Db 1590 ATTTCTGAAGTACAGCCATATTAATACTAAGTAAACTCGCCCTCGGGAACCTACACTTC 1649  
Qy 502 ArgLeuThrValThrAspSerAspGlyAlaThrAsnSerThrThrAlaAlaLeuIleVal 521  
Db 1650 AGCTTGACTGTAGTAGACTCTGTATGAGCTACCAACTCTACTCTGCAAACTGCAGTG 1709

Qy 522 AsnAsnAlaValAspTyrProProValAlaAsnAlaGlyProAsnHisThrIleThrLeu 541  
Db 1710 AACAAAGCTTGGGATTAACCCCTGTGGCCCAACGAGGCCCAACCAAGTATCACCCCTG 1769  
Qy 542 ProGlnAsnSerIleThrLeuAsnGlyAsnGlnSerSerAspAspHisGlnIleValLeu 561  
Db 1770 CCCCAAAACTCCATCACCTCTTTGGGAACCCAGAGCACTGATGATCATGGCATCACCA 1829  
Qy 562 TyrGluTrpSerLeuGlyProGlySerGluGlyLysHisValValMetGlnGlyValGln 581  
Db 1830 TATGAGTGGTCACTCAGCCCAAGCAGCAAGGGAAGTGGTGGAGATGCAGGGTGTTAGA 1889  
Qy 582 ThrProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyrThrPheGlnLeuLysVal 601  
Db 1890 ACACCAACCTTACAGCTCTCTCGATGCAAGAAGAGACTACACTTACCACCTCACAGTG 1949  
Qy 602 ThrAspSerSerArgGlnSerThrAla\*\*\*ValThrValIleValGlnProGluAsn 621  
Db 1950 ACTGACACAATAGGACAGCAGGCCACTGCTCAAGTGACTGTATTGTGCAACCTGAAAC 2009  
Qy 622 AsnArgProProValAlaValAlaGlyProAspLysGluLeuIlePheProValGluSer 641  
Db 2010 AATAAGCCTCTCTCAGGAGATGCAGGCCCAAGATAAAGAGCTGACCTTCTCTGTGGATAGC 2069  
Qy 642 AlaThrLeuAspGlySerSerSerSerAspAspHisGlyIleValPheTyrHisTrrpGlu 661  
Db 2070 ACAACCTGGATGCGAGAGAGCTCAGATGATCAGAAATATATCTCATATCTCTGGGA 2129  
Qy 662 HisValArgGlyProSerAlaVal-GluMetGluAsnIleAspLysAlaIleAlaThrVa 681  
Db 2130 AAAACACAGGACCTGATGGGTCTATCTCGAGAATGCTAACACAGCAGCGTGTACTGT 2189  
Qy 681 lThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrValLysAspGlnGlnGly 701  
Db 2190 GACTGGGTGCGCAGTGGGACCTATGTGTTCCACTGCTGCAAGATGATAGGAACCT 2249  
Qy 701 uSerSerThrSerThrLeuThrValAlaValLysLysGluAsnAsnSerProProArgAl 721  
Db 2250 GCAAAGCTGAGCTCTGTGAATGTCATTGTCGAAGACAATAAACAAACACCTATAGC 2309  
Qy 721 aArgAlaGlyArgHisValLeuValLeuProAsnAsnSerIleThrLeuAspGlySe 741  
Db 2310 CAAGATAACTGGGAATGTGGTATACCTTACCACAGCACATCAGAGCTGGATGGCTC 2369  
Qy 741 rArgSerThrAspAspGlnArgIleValSerTyrLeuTrpIleArg-AspGlyGlnSerP 761  
Db 2370 TAAGACCTCAGATGACCAAGGATAGTCAGCTACTCTGGACTCGAAAGAGGGGAGGCC 2429  
Qy 761 roAlaAlaGlyAspValIleAspGlySerAspHisSerValAlaLeuGlnLeuThrAsnL 781  
Db 2430 CAGCAGAGGGGGGTGTAAATCACTCTGACCATCACCTTATCTCTTTTCTTTCAAAACC 2489  
Qy 781 euValGluGlyValThrPheHisLeuArgValThrAspSerGlnGlyAlaSerAspT 801  
Db 2490 TGGTTGAGGNAACCTACACTTTTCACTGAAAGTGACCGATGCAAAAGGTCAGAGTGACA 2549  
Qy 801 hrAspThrAlaThrValGluValGlnProAspProArgLysSerGlyLeuValGluLeu 821  
Db 2550 CAGACCGACCACTGTGGAGGTGAACCTGATCCAGGAAAAACAACCTGGTGGAGATCA 2609  
Qy 821 hrLeuGlnValGlyValGlyLysLeuThrGluGlnArgLysAspThrLeuValArgGlnL 841  
Db 2610 TCTTGGATATCAACGTCAGTCAGCTAAGTGAAGGGGTGAAGGGGATGTTCTCGCCAGA 2669  
Qy 841 euAlaValLeuLeuAsnValLeuAspSerAspIleLysValGlnLysIleArgAlaHis 861  
Db 2670 TTGGGGTCTCTCTGGGGTGTGATTTCCGACATCATTTGTGCAAAAGATTTCAGCCGTACA 2729  
Qy 861 erAspLeuSerThrValIleValPheTyrValGlnSerArgProPheLysValLeuL 881  
Db 2730 CGGAGCAGAGCAGCAAAATGGTATTTTTTTGTTCAAAACGAGCCCTCCCAACCATCTTTA 2789  
Qy 881 ySAlaAlaGluValAlaArgAsnLeuHisMetArgLeuSerLysGluLysAlaAspPheL 901







Db 543 GGCCCCAACCAAGTGATCACCTGCCCCCAAACTCCATCACCTCTTTGGGAACAGAGC 602  
 Qy 554 SerAspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyLys 573  
 Db 603 ACTGATGATCATGGCATCACAGCTATGAGTGGTCACTCAGCCCCAAGCAGCAAGGGGAAA 662  
 Qy 574 HisValValMetGlnGlyValGlnThrProTyrLeuHisLeuSerAlaMetGlnGluGly 593  
 Db 663 GTGTGGGATGCAAGGGTTAGAACCAACCACTTACAGCTCTCTGCCGATGCAGAAGGA 722  
 Qy 594 AspTyrThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla\*\*Val 613  
 Db 723 GACTACACTTACCAGCTCACAGTGACTGACACATAGGACAGCAGGCCCTGCTCAAGTG 782  
 Qy 614 ThrValIleValGlnProGluAsnAsnArgProProValAlaValAlaGlyProAspLys 633  
 Db 783 ACTGTATTGTGCAACCTTGAACCAATTAAGCTCTCTCAGGCAGATGTCAGGCCCAAGATAA 842  
 Qy 634 GluLeuIlePheProValGluSerAlaThrLeuAspGlySerSerSerAspAspHis 653  
 Db 843 GAGCTGACCTTCTGTGGATAGCACCAACCTCTGGATGGCAGCAAGACTCAGATGATCAG 902  
 Qy 654 GlyIleValPheTyrHisTrpGluHisValArgGlyProSerAlaValGluMetGluAsn 673  
 Db 903 AAAATTATCATATCTCTGGGAAAAACACAGGGACCTGTAGGGTGCACTCGAGAT 962  
 Qy 674 IleAspLysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeu 693  
 Db 963 GCTAACAGCAGCTGTGTACTGTGACTGGGCTGCAAGTGGGAGCTATGTGTTCACCTTG 1022  
 Qy 694 ThrValLysAspGlnGlnGlyLeuSerSerThrThrLeuThrValAlaValLysLys 713  
 Db 1023 ACTGTCAAAAGTAGAGGAACCTGCAAGCCAGAGCTCTGTGAATGTCAATGTCAAAGAA 1082  
 Qy 714 GluAsnAsnSerProProArgAlaAlaGlyArgHisValLeuValLeuProAsn 733  
 Db 1083 GAATATAACAACCACTATAGCCCAAGATACTGGGAATGTGGTGAATTACCTTACCACG 1142  
 Qy 734 AsnSerIleThrLeuAspGlySerArgSerThrAspGlnArgIleValSerTyrIleu 753  
 Db 1143 AGCACAGCAGCTGGATGGCTTAAGTCTCAGATGACAGGAAGTAAGTCACTACCTC 1202  
 Qy 754 TrpIleArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSer 773  
 Db 1203 TGGACTCGAGATAGGGGAGGCCAGCAGCAGGGAGGTGTAAATCACTCTGACCATCAC 1262  
 Qy 774 ValAlaLeuGlnLeuThrAsnLeuValGluGlyValTyrThrPheHisLeuArgValThr 793  
 Db 1263 CCTATCTTTTCTTTCAAACTGTTGAGGGAACCTACACTTTTTCACCTGAAAGTGACC 1322  
 Qy 794 AspSerGlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArg 813  
 Db 1323 GATCAAAAGGTGAGAGTGACACAGCCGACCACTGTGGAGGTGAAACCTGATCCACG 1382  
 Qy 814 LysSerGlyLeuValGlnLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArg 833  
 Db 1383 AAAAAACAACCTGGTGGAGATCATCTTGGATATCAACGTCAGTCAGTAACCTGAGAGGCTG 1442  
 Qy 834 LysAspThrLeuValArgGlnLeuAlaValLeuLeuAsnValLeuAspSerAspIleLys 853  
 Db 1443 AAGGGATGTTTATCCCGCAATGTTGGGTCTCTCTGGGGGTGGTGGATCCGATCATTT 1502  
 Qy 854 ValGlnLysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSer 873  
 Db 1503 GTGCAAAAGATTACCGCTACACGAGCAGCAGCAACCAAAATGGTATTTTTTGTTCAAAAC 1562  
 Qy 874 ArgProProPheLysValLeuLysAlaAlaGluValAlaArgAsnLeuHisMetArgLeu 893  
 Db 1563 GAGCCTCCCAACCATGCTTCAAAGGCCATGAGGTGGCAGCGATGCTCAAGAGTGAGCTG 1622  
 Qy 894 SerLysGluLysAlaAspPheLeuPheLysValLeuArgValAspThrAlaGlyCys 913  
 Db 1623 CGGAAGCAAAAGGCAGACTTTTTTGTATTTACAGAGCTTGGGAAGTCAACACTGTCACTGT 1682

Qy 914 LeuLeuLysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCysIleCysSer 933  
 Db 1683 CAGCTGAACCTTCCGACCACTGTGACTCGTTTCAACCAACGCTGATCTGTGAC 1742  
 Qy 934 HisLeuTrpMetGluAsnLeuIleGlnArgTyrIleTrpAspGlyGluSerAsnCysGlu 953  
 Db 1743 CCTTTTGGATGGAGAAATTCATCAAGTGCAGCTGAGGGATGGAGACGACACTGTGAG 1802  
 Qy 954 TrpSerIlePheTyrValThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPhe 973  
 Db 1803 TGGAGCGTGTATATGTTATCATTTGCTACCTTTGTCAATTTGTTGCTTGGGAATCCTG 1862  
 Qy 974 ThrTrpLeuCysIleCysCysCysArgGlnLysArgThrLysIleArgLysLysThr 993  
 Db 1863 TCTTTGGACTGTGATCTGTGTTGTAAGAGGCAAAA---GGAAAAACCAAGAGGAAAGC 1919  
 Qy 994 LysTyrThrIleLeuAspAsnMetAspGluGlnGluArgMetGluLeuArgPro----- 1011  
 Db 1920 AAGTACAGATCTCTGGATGCCACGAT---CAGGAAGCCTGGAGCTGAGCCAACTCC 1976  
 Qy 1012 LysTyrGlyIleLysHisArgSerThrGluHisAsnSerSerLeuMetValSerGluSer 1031  
 Db 1977 CGACAGGCATCAAAACAGAAAGGCTTTTGTCTAAGTAGCAGCTGATGCACTCCGAGTCA 2036  
 Qy 1032 GluPheAspSerAspGlnAspThrIlePheSerArgGluLysMetGluArgGlyAsnPro 1051  
 Db 2037 GAGCTGGACACGCGAT---GATGCCATCTTTATACATGGCCAGACCGCAGAGAGGCAAACTC 2093  
 Qy 1052 LysValSerMetAsnGlySerIleArgAsnGly 1062  
 Db 2094 CTGCATGTCGAATGGCTCTGTACCCCAACGGG 2126

## RESULT 15

AAV58761  
 ID AAV58761 standard; cDNA; 2501 BP.  
 XX  
 AC AAV58761;  
 DT  
 DT 18-JAN-1999 (first entry)  
 DE Human secreted protein cw1233\_3 cDNA.  
 XX  
 KW Secreted protein; human; cw1233\_3; db.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 364..780  
 FT /\*tag= a  
 XX  
 XX WO9840404-A2.  
 PD 17-SEP-1998.  
 XX  
 XX 09-MAR-1998; 98WO-US004601.  
 PF  
 XX 11-MAR-1997; 97US-00815381.  
 PR 06-MAR-1998; 98US-00036321.  
 XX  
 XX (GEMY ) GENETICS INST INC.  
 XX  
 XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;  
 PI Spaulding V, Agostino MJ;  
 XX  
 DR WPI; 1998-520802/44.  
 DR P-PSDB; AAW69431.  
 XX  
 XX New isolated polynucleotides and secreted proteins - useful as, e.g.  
 PT nutritional additives, immunostimulators, haemopoiesis regulators and  
 PT as diagnostic agents.  
 XX  
 PS Claim 32; Page 88-89; 119pp; English.

XX This full-length cDNA clone, designated cw1233.3, codes for a novel human  
 CC secreted protein (see AAW69430), i.e. cw1233.3 protein. It was isolated  
 CC from a human foetal brain cDNA library using methods which are selective  
 CC for cDNAs encoding secreted proteins, or was identified as encoding a  
 CC secreted or transmembrane protein on the basis of computer analysis of  
 CC the encoding protein. Homology is shown to some database sequences. The  
 CC invention provides isolated polynucleotides (see AAV58754-63) obtained  
 CC from human adult testis, brain, retina or placenta, or from foetal kidney  
 CC or brain cDNA libraries. These are all deposited at ATCC 98353. They  
 CC encode novel human secreted proteins (see AAW69423-33) that may have e.g.  
 CC nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, antiinflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC or other activities. They may also be used for diagnostic purposes  
 XX  
 SQ Sequence 2501 BP; 682 A; 649 C; 646 G; 523 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 1,06e-100 Length: 2501  
 Score: 1817.50 Matches: 366  
 Percent Similarity: 73.23% Conservative: 80  
 Best Local Similarity: 60.10% Mismatches: 158  
 Query Match: 32.57% Indels: 7  
 DB: 2 Gaps: 4

US-10-764-390-3 (1-1072) x AAV58761 (1-2501)

QY 456 GlySerGlnSerThrAspAspThrGluLeuValSerTyrHisTrpGluGluLeuAenGly 475  
 DB 1 GCGAGTCAAGCACTGATGATGATAAATCGTTTCAGTACCATTTGGAGAGCACTTAAGGGG 60  
 QY 476 PropheileGluGluLysThrSerValAspSerProValLeuArgLeuSerAenLeuAsp 495  
 DB 61 CCTTAAGAGAGAGAGATTCTTGAAGATACAGCCATATAAATTAAGTAAGTAACTCGTC 120  
 QY 496 ProGlyAsnTyrSerPheArgLeuThrValThrAspSerAspGlyAlaThrAenSerThr 515  
 DB 121 CCTGGGAACCTACACTTTCAGCTTGACTGTAGTACTCTGATGGAGCTACCAACTCTACT 180  
 QY 516 ThrAlaAlaLeuLeuValAenAenAlaValAspTyrProProValAlaAenAlaGlyPro 535  
 DB 181 ACTGCAAACTGACAGTGAACAAAGCTGTGGATTACCCCTCTGTGGCCAAACGAGGCCCC 240  
 QY 536 AsnHisThrIleThrLeuProGlnAenSerIleThrLeuAenGlyAenGlnSerSerAsp 555  
 DB 241 AACCAAGTGATCACCTGCCCCCAAACTCCATCACCTCTTTGGGAACCCAGAGCACTGAT 300  
 QY 556 AspHisGlnIleValLeuTyrGluTrpSerLeuGlyProGlySerGluGlyLysHisVal 575  
 DB 301 GATCATGGGATCACCACTGATGAGTGTGCTCACTCAGCCCAAGCAGCAAGGGAAGTGGTG 360  
 QY 576 ValMetGlnGlyValGlnProTyrLeuHisLeuSerAlaMetGlnGluGlyAspTyr 595  
 DB 361 GAGATCAGGGGTGTAGAACCACTTACAGCTCTCTGCGATGCAAGAGAGAGACTAC 420  
 QY 596 ThrPheGlnLeuLysValThrAspSerSerArgGlnGlnSerThrAla\*\*\*ValThrVal 615  
 DB 421 ACTTACCAGCTCACAGTGTGACACAATAGGACAGCAGGCCACTGCTCAAGTACTGTT 480  
 QY 616 IleValGlnProGlnAenAenArgProProValAlaValAlaGlyProAspLysGluLeu 635  
 DB 481 ATTGTGCAACCTGAAACCAATAAGCTCTCTCAGGCAAGATGCGGCCCTCAGATAAAGAGCTG 540  
 QY 636 IlePheProValGluSerAlaThrLeuAspGlySerSerSerAspAspHisGlyIle 655  
 DB 541 ACCCTTCTCTGTGTAGACCAACCTTGGATGGGCAAGCAAGCTCAGATGATCAGAAAT 600  
 QY 656 ValPheTyrHisTrpGluHisValArgGlyProSerAlaValGluMetGluAenIleAsp 675  
 DB 601 ATCTCATATCTCTGGGAAAAAACAACAGGGGACCTGATGGGGTGCAGCTCGAGATGCTAAC 660

QY 676 LysAlaIleAlaThrValThrGlyLeuGlnValGlyThrTyrHisPheArgLeuThrVal 695  
 DB 661 AGCAGTGTGTACTGTGACTGGGCTGCAAGTGGGACCTATGTGTTCACCTTGACTGTC 720  
 QY 696 LysAspGlnGlnGlyLeuSerSerThrSerThrLeuThrValAlaValLysLysGluAen 715  
 DB 721 AAAGATGAGAGGAACCTGCAAAAGCCAGAGCTCTGTGAATGTCAATGTCAAAAGAGAATA- 779  
 QY 716 AsnSerProProArgAlaArgAlaGlyArgHisValLeuValLeuProAsnAenSer 735  
 DB 780 AACAAACACCATATAGCCAGATTAACCTGGGAATGTGGTGTATACCTACCCACGAGACA 839  
 QY 736 IleThrLeuAspGlySerArgSerThrAspAspGlnArgIleValSerTyrLeuTrpIle 755  
 DB 840 CGAGAGCTGGATGGCTTAAGTCTTCAGATGACCAAGGAATAGTCAGTACCTCTGGACT 899  
 QY 756 ArgAspGlyGlnSerProAlaAlaGlyAspValIleAspGlySerAspHisSerValAla 775  
 DB 900 CGAGATGAGGGGAGCCACGAGCAGGGGGGTGTTAAATCACTCTGACCATCACCCCTATC 959  
 QY 776 LeuGlnLeuThrAenLeuValGluGlyValTyrThrPheHisLeuArgValThrAspSer 795  
 DB 960 CTTTTTCTTTCAAACCTGGTTGAGGGAAACCTACACTTTTTCACCTGAAAGTGAACCATGCA 1019  
 QY 796 GlnGlyAlaSerAspThrAspThrAlaThrValGluValGlnProAspProArgLysSer 815  
 DB 1020 AAGGTGAGAGTGACACAGACCGGACCACTGTGGGGTGAACCTGATCCCGAGGAAGAAC 1079  
 QY 816 GlyLeuValGluLeuThrLeuGlnValGlyValGlyGlnLeuThrGluGlnArgLysAsp 835  
 DB 1080 AACCTGGTGGAGATCATCTTGGATATCAACGTCAGTCACTAAGTGAAGTGAAGGGG 1139  
 QY 836 ThrLeuValArgGlnLeuAlaValLeuAenValLeuAenValLeuAenSerAspIleLysValGln 855  
 DB 1140 ATGTTTATCCCGCAGATTGGGGTCTCTCTGGGGTGTGGATTCGACATCATTTGTGCAA 1199  
 QY 856 LysIleArgAlaHisSerAspLeuSerThrValIleValPheTyrValGlnSerArgPro 875  
 DB 1200 AAGATTACGCGTACACGAGCAGACGACCAAAATGGTATTTTTTGTTCNAACGAGCT 1259  
 QY 876 PropheLysValLeuLysAlaAlaGluValAlaArgAenLeuHisMetArgLeuSerLys 895  
 DB 1260 CCCCACCAAGCTTCAAAAGGCCATGAGTGGCAGCGATGCTCAAGAGTGAAGTGGGAAG 1319  
 QY 896 GluLysAlaAspPheLeuLeuPheLysValLeuArgValAspThrAlaGlyCysLeuLeu 915  
 DB 1320 CAAAGGCGAGACTTTTGTATTCAGAGCTTGGGAAGTCAACACTGTCAATGTCAAGCTG 1379  
 QY 916 LysCysSerGlyHisGlyHisCysAspProLeuThrLysArgCysIleCysSerHisLeu 935  
 DB 1380 AACTGTTCCGACCATGGTACTGTGACTCGTTCACCAAAACGCTGTATCTGTGACCCCTTT 1439  
 QY 936 TrpMetGluAenLeuIleGlnArgTyrIleTrpAspGlyGluSerAenCysGluTrpSer 955  
 DB 1440 TGGATGGAGAATTTTCATCAAGGTGAGCTGAGGGATGGAGACAGCAACTGTGAGTGGAGC 1499  
 QY 956 IlePheTyrValThrValLeuAlaPheThrLeuIleValLeuThrGlyGlyPheThrTrp 975  
 DB 1500 GTGTATATGTATCATTTGCTACCTTTGTCTATTTGTGTGTGGTTCCTGGGAATCTGTCTGG 1559  
 QY 976 LeuCysIleCysCysLysArgGlnLysArgThrLysIleArgLysLysThrLysTyr 995  
 DB 1560 ACTGTGATCTGTGTGTGTAAAGGCGAAAAA---GGAAAAACCCAGAGGAAGAAAGATAC 1616  
 QY 996 ThrIleLeuAspAenMetAspGluGlnArgMetGluLeuArgPro-----LysTyr 1013  
 DB 1617 AAGAT-CTGGATGCCACGGAT---CAGGAAAGCTTGGAGCTGAAGCCAACTCCCGAGCA 1672  
 QY 1014 GlyIleLysHisArgSerThrGluHisAenSerSerLeuMetValSerGluSerGluPhe 1033  
 DB 1673 GGCATCAACAGAAAAAGGCGCTTTTGTAAAGTAGCAGCTGATGCACCTCCGAGTCAGAGCTG 1732





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:22:50 ; Search time 70 Seconds  
(without alignments)  
6370.973 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 1072  
Sequence: 1 MAPPTGVLSLLLVTTIAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1854112 seqs, 416015017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 458990

Minimum DB seq length: 0

Maximum DB seq length: 35

Post-processing: Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US11A\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	1.7	19	18	US-10-450-763-52425
2	16	1.5	16	16	US-10-764-390-43
3	14	1.3	24	16	US-10-764-390-265
4	14	1.3	29	16	US-10-764-390-262
5	14	1.3	29	16	US-10-764-390-268
6	9	0.8	9	16	US-10-764-390-44
7	9	0.8	9	16	US-10-764-390-45
8	9	0.8	9	16	US-10-764-390-46
9	9	0.8	9	16	US-10-764-390-47
10	9	0.8	9	16	US-10-764-390-48
11	9	0.8	9	16	US-10-764-390-49

12	9	0.8	9	16	US-10-764-390-50	Sequence 50, Appl
13	9	0.8	9	16	US-10-764-390-51	Sequence 51, Appl
14	9	0.8	9	16	US-10-764-390-52	Sequence 52, Appl
15	9	0.8	9	16	US-10-764-390-53	Sequence 53, Appl
16	9	0.8	9	16	US-10-764-390-54	Sequence 54, Appl
17	9	0.8	9	16	US-10-764-390-55	Sequence 55, Appl
18	9	0.8	9	16	US-10-764-390-56	Sequence 56, Appl
19	9	0.8	9	16	US-10-764-390-59	Sequence 59, Appl
20	9	0.8	9	16	US-10-764-390-60	Sequence 60, Appl
21	9	0.8	9	16	US-10-764-390-61	Sequence 61, Appl
22	9	0.8	9	16	US-10-764-390-62	Sequence 62, Appl
23	9	0.8	9	16	US-10-764-390-63	Sequence 63, Appl
24	9	0.8	9	16	US-10-764-390-64	Sequence 64, Appl
25	9	0.8	9	16	US-10-764-390-65	Sequence 65, Appl
26	9	0.8	9	16	US-10-764-390-66	Sequence 66, Appl
27	9	0.8	9	16	US-10-764-390-67	Sequence 67, Appl
28	9	0.8	9	16	US-10-764-390-68	Sequence 68, Appl
29	9	0.8	9	16	US-10-764-390-69	Sequence 69, Appl
30	9	0.8	9	16	US-10-764-390-70	Sequence 70, Appl
31	9	0.8	9	16	US-10-764-390-71	Sequence 71, Appl
32	9	0.8	9	16	US-10-764-390-72	Sequence 72, Appl
33	9	0.8	9	16	US-10-764-390-73	Sequence 73, Appl
34	9	0.8	9	16	US-10-764-390-74	Sequence 74, Appl
35	9	0.8	9	16	US-10-764-390-75	Sequence 75, Appl
36	9	0.8	9	16	US-10-764-390-76	Sequence 76, Appl
37	9	0.8	9	16	US-10-764-390-77	Sequence 77, Appl
38	9	0.8	9	16	US-10-764-390-78	Sequence 78, Appl
39	9	0.8	9	16	US-10-764-390-79	Sequence 79, Appl
40	9	0.8	9	16	US-10-764-390-80	Sequence 80, Appl
41	9	0.8	9	16	US-10-764-390-81	Sequence 81, Appl
42	9	0.8	9	16	US-10-764-390-82	Sequence 82, Appl
43	9	0.8	9	16	US-10-764-390-83	Sequence 83, Appl
44	9	0.8	9	16	US-10-764-390-84	Sequence 84, Appl
45	9	0.8	9	16	US-10-764-390-85	Sequence 85, Appl

ALIGNMENTS

RESULT 1

US-10-450-763-52425  
; Sequence 52425, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 52425  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(19)  
; OTHER INFORMATION: xaa = X or \* as defined in Table 2  
US-10-450-763-52425

Query Match 1.7%; Score 18; DB 18; Length 19;  
Best Local Similarity 100.0%; Pred.No. 4.1e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MAPPTGVLSLLLVTTIA 18  
|||||||

```
Db      1  MAPPTGVLSLLLVITIA 18

RESULT 2
US-10-764-390-43
; Sequence 43, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-43

Query Match      1.5%; Score 16; DB 16; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      522  NNAVDYPPVANAGPNH 537
Db      1  NNAVDYPPVANAGPNH 16

RESULT 3
US-10-764-390-265
; Sequence 265, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-265

Query Match      1.3%; Score 14; DB 16; Length 24;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      20  CARQKQCEGRYTSN 33
        11  CARQKQCEGRYTSN 24

RESULT 4
US-10-764-390-262
; Sequence 262, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 262
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-262

Query Match      1.3%; Score 14; DB 16; Length 29;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      143  FLGRDNGLEEMSEY 156
Db      1  FLGRDNGLEEMSEY 14

RESULT 5
US-10-764-390-268
; Sequence 268, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 268
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-268
```

```

Db      1  GVLSSLLLL 9

RESULT 8
US-10-764-390-46
; Sequence 46, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1D6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-46

Query Match      0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy      22  RKQCSSEGR 30
      |||||
Db      1  RKQCSSEGR 9

RESULT 9
US-10-764-390-47
; Sequence 47, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1D6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-47

Query Match      0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy      22  TGVLSLLLL 13
      |||||
Db      1  TGVLSLLLL 9

RESULT 6
US-10-764-390-44
; Sequence 44, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1D6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-44

Query Match      0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5  TGVLSLLLL 13
      |||||
Db      1  TGVLSLLLL 9

RESULT 7
US-10-764-390-45
; Sequence 45, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1D6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-45

```

```
; SEQ ID NO 47
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-47

Query Match          0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GRTYSNAVI 36
   |||||
Db 1 GRTYSNAVI 9

RESULT 10
US-10-764-390-48
; Sequence 48, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1d6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-48

Query Match          0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 NAVISPNLE 41
   |||||
Db 1 NAVISPNLE 9

RESULT 11
US-10-764-390-49
; Sequence 49, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1d6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
```

```
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-49

Query Match          0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 IMRVSHTPP 53
   |||||
Db 1 IMRVSHTPP 9

RESULT 12
US-10-764-390-50
; Sequence 50, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1d6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-50

Query Match          0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 CCCLSSCDL 69
   |||||
Db 1 CCCLSSCDL 9

RESULT 13
US-10-764-390-51
; Sequence 51, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254p1d6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
```



```
; CURRENT APPLICATION NUMBER: US/10/764,390
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-51

Query Match      0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      62 CDLSGCCDLA 70
Db      1 CDLSGCCDLA 9

RESULT 14
US-10-764-390-52
; Sequence 52, Application US/10/764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-52

Query Match      0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      77 CYLVSCPHK 85
Db      1 CYLVSCPHK 9

RESULT 15
US-10-764-390-53
; Sequence 53, Application US/10/764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-53

Query Match      0.8%; Score 9; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      94 GPIRSYLTFF 102
Db      1 GPIRSYLTFF 9

Search completed: October 12, 2005, 10:27:51
Job time : 70 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:21:44 ; Search time 26 seconds  
(without alignments)  
3077.838 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 1072  
Sequence: 1 MAPPTGVLSLLLVTTIAGC.....VSMNGSIRNGASFYCSKDR 1072

Scoring table: OIIGO  
Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 250370

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Listing first 45 summaries

Database : 1: Issued Patents AA.\*  
2: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pap.\*  
7: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	0.6	7	2	US-08-666-473-53
2	6	0.6	10	3	US-09-001-984C-78
3	6	0.6	10	4	US-09-396-347F-78
4	6	0.6	11	3	US-09-261-855-14
5	6	0.6	11	4	US-09-873-637-14
6	6	0.6	12	2	US-08-096-762-205
7	6	0.6	12	3	US-09-042-353-320
8	6	0.6	12	3	US-08-758-417A-168
9	6	0.6	15	3	US-09-044-718-8
10	6	0.6	15	4	US-10-062-848-8
11	6	0.6	16	4	US-08-537-871A-41
12	6	0.6	16	4	US-08-537-871A-45
13	6	0.6	17	1	US-08-347-198A-11
14	6	0.6	17	2	US-08-248-839C-53
15	6	0.6	17	3	US-08-333-844A-35
16	6	0.6	17	3	US-08-335-844A-50
17	6	0.6	17	4	US-09-129-366-35
18	6	0.6	17	4	US-09-129-366-50
19	6	0.6	18	2	US-08-637-759B-185
20	6	0.6	18	2	US-08-280-864A-6
21	6	0.6	18	3	US-08-871-355A-185
22	6	0.6	18	3	US-09-092-291-6
23	6	0.6	18	3	US-09-201-945-185
24	6	0.6	19	4	US-09-355-040-27
25	6	0.6	20	2	US-08-053-451B-122
26	6	0.6	20	4	US-09-053-611-17
27	6	0.6	20	4	US-09-889-136-6

28	6	0.6	20	4	US-09-889-136-11	Sequence 11, Appl
29	6	0.6	22	1	US-08-483-146A-6	Sequence 6, Appl
30	6	0.6	22	1	US-08-232-513A-15	Sequence 15, Appl
31	6	0.6	22	1	US-08-484-594A-6	Sequence 6, Appl
32	6	0.6	22	3	US-09-231-159-4	Sequence 4, Appl
33	6	0.6	22	3	US-08-611-307-4	Sequence 4, Appl
34	6	0.6	22	4	US-09-076-258A-6	Sequence 6, Appl
35	6	0.6	24	1	US-08-628-291-14	Sequence 14, Appl
36	6	0.6	24	2	US-09-128-722-14	Sequence 14, Appl
37	6	0.6	25	4	US-09-205-258-617	Sequence 617, App
38	6	0.6	26	3	US-08-467-023-229	Sequence 229, App
39	6	0.6	26	3	US-08-467-023-232	Sequence 232, App
40	6	0.6	27	4	US-09-471-276-1290	Sequence 1290, Ap
41	6	0.6	27	4	US-09-936-885A-13	Sequence 13, Appl
42	6	0.6	28	4	US-09-181-341-10	Sequence 10, Appl
43	6	0.6	29	3	US-09-348-578-4	Sequence 3, Appl
44	6	0.6	29	3	US-09-023-731-3	Sequence 3, Appl
45	6	0.6	29	4	US-09-699-684-4	Sequence 4, Appl

## ALIGNMENTS

### RESULT 1

US-08-666-473-53  
; Sequence 53, Application US/08666473  
; Patent No. 5843713  
; GENERAL INFORMATION:  
; APPLICANT: YOSHIDA, Aruto  
; APPLICANT: TAKEUCHI, Makoto  
; TITLE OF INVENTION: PEPTIDE SEQUENCE THAT FORMS MUCIN SUGAR  
; TITLE OF INVENTION: CHAIN AND TECHNIQUE FOR MODIFYING PROTEIN TO BE LINKED  
; TITLE OF INVENTION: WITH MUCIN SUGAR CHAIN  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/666,473  
; FILING DATE: 19-SEP-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP95/02238  
; FILING DATE: 01-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-22101  
; FILING DATE: 09-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-269111  
; FILING DATE: 01-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 16887/837  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-666-473-53

Query Match 0.6%; Score 6; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 ASTPAP 227  
Db 2 ASTPAP 7

## RESULT 2

US-09-001-984C-78  
; Sequence 78, Application US/09001984C  
; Patent No. 6245331  
; GENERAL INFORMATION:  
; APPLICANT: Laal, Suman  
; APPLICANT: Zolla-Pazner, Susan  
; APPLICANT: Belisle, John T  
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE  
; FILE REFERENCE: NYU-011  
; CURRENT APPLICATION NUMBER: US/09/001,984C  
; CURRENT FILING DATE: 1997-12-31  
; PRIOR APPLICATION NUMBER: 60/034,003  
; PRIOR FILING DATE: 1996-12-31  
; NUMBER OF SEQ ID NOS: 106  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 78  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis strain H37Rv  
US-09-001-984C-78

Query Match 0.6%; Score 6; DB 3; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 APAPPV 364  
Db 3 APAPPV 8

## RESULT 3

US-09-396-347F-78  
; Sequence 78, Application US/09396347F  
; Patent No. 6506384  
; GENERAL INFORMATION:  
; APPLICANT: Laal, Suman  
; APPLICANT: Zolla-Pazner, Susan  
; APPLICANT: Belisle, John T  
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE  
; FILE REFERENCE: 32004-189276  
; CURRENT APPLICATION NUMBER: US/09/396,347F  
; CURRENT FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: 09/001,984  
; PRIOR FILING DATE: 1997-12-31  
; NUMBER OF SEQ ID NOS: 106  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 78  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis strain H37Rv  
US-09-396-347F-78

Query Match 0.6%; Score 6; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 APAPPV 364  
Db 3 APAPPV 8

## RESULT 4

US-09-261-855-14  
; Sequence 14, Application US/09261855A  
; Patent No. 6255055  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Jeffrey  
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN  
; FILE REFERENCE: 960296.95131  
; CURRENT APPLICATION NUMBER: US/09/261,855A  
; CURRENT FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-261-855-14

Query Match 0.6%; Score 6; DB 3; Length 11;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 442 LQELTL 447  
Db 5 LQELTL 10

## RESULT 5

US-09-873-637-14  
; Sequence 14, Application US/09873637  
; Patent No. 6794151  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Jeffrey  
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN  
; FILE REFERENCE: 960296.95131  
; CURRENT APPLICATION NUMBER: US/09/873,637  
; CURRENT FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-873-637-14

Query Match 0.6%; Score 6; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 442 LQELTL 447  
Db 5 LQELTL 10

## RESULT 6

US-08-096-762-205  
; Sequence 205, Application US/08096762  
; Patent No. 5814318  
; GENERAL INFORMATION:  
; APPLICANT: Lonberg, Nils  
; APPLICANT: Kay, Robert M.  
; TITLE OF INVENTION: Transgenic No. 5814318-Human Animals for  
; NUMBER OF SEQUENCES: 210  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: One Market Plaza, Steuart Tower, Suite 200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/904,068  
FILING DATE: 23-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/853,408  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,279  
FILING DATE: 17-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William W.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 14643-9-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 205:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-096-762-205

Query Match 0.6%; Score 6; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VTVSSE 412  
|||||  
Db 7 VTVSSE 12

RESULT 7  
US-09-042-353-320  
Sequence 320, Application US/09042353  
Patent No 6255458  
GENERAL INFORMATION:  
APPLICANT: Lonberg, Nils  
APPLICANT: Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for  
TITLE OF INVENTION: Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 421  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/042,353  
FILING DATE: 13-MAR-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,279  
FILING DATE: 17-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/853,408  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/904,068  
FILING DATE: 23-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/155,301  
FILING DATE: 18-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/165,699  
FILING DATE: 10-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/209,741  
FILING DATE: 09-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/352,322  
FILING DATE: 07-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/544,404  
FILING DATE: 10-OCT-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US96/16433  
FILING DATE: 10-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/758,417  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/21803  
FILING DATE: 01-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 014643-00904005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 320:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-042-353-320

Query Match 0.6%; Score 6; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VTVSSE 412  
|||||

Db 7 VTVSSE 12

RESULT 8  
US-08-758-417A-168  
; Sequence 168, Application US/08758417A  
; Patent No. 6300129  
; GENERAL INFORMATION:  
; APPLICANT: Lonberg, Nils  
; Kay, Robert M.  
; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for  
; Producing Heterologous Antibodies  
; NUMBER OF SEQUENCES: 417  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/758,417A  
; FILING DATE: 02-Dec-1996  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/728,463  
; FILING DATE: 10-OCT-1996  
; APPLICATION NUMBER: US 08/544,404  
; FILING DATE: 10-OCT-1995  
; APPLICATION NUMBER: US 08/352,322  
; FILING DATE: 07-DEC-1994  
; APPLICATION NUMBER: US 08/209,741  
; FILING DATE: 09-MAR-1994  
; APPLICATION NUMBER: US 08/165,699  
; FILING DATE: 10-DEC-1993  
; APPLICATION NUMBER: US 08/161,739  
; FILING DATE: 03-DEC-1993  
; APPLICATION NUMBER: US 08/155,301  
; FILING DATE: 18-NOV-1993  
; APPLICATION NUMBER: US 08/096,762  
; FILING DATE: 22-JUL-1993  
; APPLICATION NUMBER: US 08/053,131  
; FILING DATE: 26-APR-1993  
; APPLICATION NUMBER: US 07/990,860  
; FILING DATE: 16-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Serafini, Andrew T.  
; REGISTRATION NUMBER: 41,303  
; REFERENCE/DOCKET NUMBER: 014643-0090300S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 168:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 168:  
US-08-758-417A-168

Query Match 0.6%; Score 6; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VTVSSE 412  
|||||

Db 7 VTVSSE 12

RESULT 9  
US-09-044-718-8  
; Sequence 8, Application US/09044718  
; Patent No. 6391605  
; GENERAL INFORMATION:  
; APPLICANT: KOSTREWA, Dirk  
; PASAMONTES, Luis  
; APPLICANT: TOMSCHY, Andrea  
; APPLICANT: van LOON, Adolphus  
; APPLICANT: VOGEL, Kurt  
; APPLICANT: WYSS, Markus  
; TITLE OF INVENTION: MODIFIED PHYTASES  
; FILE REFERENCE: Modified Phytases  
; CURRENT APPLICATION NUMBER: US/09/044,718  
; CURRENT FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: EP 97810175.6  
; PRIOR FILING DATE: 1997-03-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Talaromyces thermophilus  
US-09-044-718-8

Query Match 0.6%; Score 6; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SLLLIIV 15  
|||||  
Db 2 SLLLIIV 7

RESULT 10  
US-10-062-848-8  
; Sequence 8, Application US/10062848  
; Patent No. 6734004  
; GENERAL INFORMATION:  
; APPLICANT: KOSTREWA, Dirk  
; PASAMONTES, Luis  
; APPLICANT: TOMSCHY, Andrea  
; APPLICANT: van LOON, Adolphus  
; APPLICANT: VOGEL, Kurt  
; APPLICANT: WYSS, Markus  
; TITLE OF INVENTION: MODIFIED PHYTASES  
; FILE REFERENCE: Modified Phytases  
; CURRENT APPLICATION NUMBER: US/10/062,848  
; CURRENT FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 09/044,718  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: EP 97810175.6  
; PRIOR FILING DATE: 1997-03-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Talaromyces thermophilus  
US-10-062-848-8

Query Match 0.6%; Score 6; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SLLLIIV 15  
|||||  
Db 2 SLLLIIV 7



## RESULT 11

US-08-537-871A-41  
; Sequence 41, Application US/08537871A  
; Patent No. 6838254  
; GENERAL INFORMATION:  
; APPLICANT: Raymond HAMERS  
; APPLICANT: Cecile HAMERS-CASERMAN  
; APPLICANT: Serge V. M. MUYLDERMANS  
; APPLICANT: Leon G. J. FRENKEN  
; APPLICANT: Cornelis T. VERRIPS  
; TITLE OF INVENTION: Production of antibodies or (functionalized)  
; fragments thereof derived from heavy chain immunoglobulins  
; of Camelidae  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 Mb disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,871A  
; FILING DATE: 29-JAN-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/01442  
; FILING DATE: 28-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EPO 93202079.5  
; FILING DATE: 15-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EPO 93201454.1  
; FILING DATE: 19-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EPO 93201239.6  
; FILING DATE: 29-APR-1993  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: See figure 6

US-08-537-871A-41

Query Match 0.6%; Score 6; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 407 VTSVSE 412

Db 1 VTSVSE 6

## RESULT 12

US-08-537-871A-45  
; Sequence 45, Application US/08537871A  
; Patent No. 6838254  
; GENERAL INFORMATION:  
; APPLICANT: Raymond HAMERS  
; APPLICANT: Cecile HAMERS-CASERMAN  
; APPLICANT: Serge V. M. MUYLDERMANS  
; APPLICANT: Leon G. J. FRENKEN  
; APPLICANT: Cornelis T. VERRIPS  
; TITLE OF INVENTION: Production of antibodies or (functionalized)  
; fragments thereof derived from heavy chain immunoglobulins

; TITLE OF INVENTION: of Camelidae  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 Mb disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,871A  
; FILING DATE: 29-JAN-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/01442  
; FILING DATE: 28-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EPO 93202079.5  
; FILING DATE: 15-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EPO 93201454.1  
; FILING DATE: 19-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EPO 93201239.6  
; FILING DATE: 29-APR-1993  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: See figure 19

US-08-537-871A-45

Query Match 0.6%; Score 6; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 407 VTSVSE 412

Db 1 VTSVSE 6

## RESULT 13

US-08-347-198A-11  
; Sequence 11, Application US/08347198A  
; Patent No. 5747046  
; GENERAL INFORMATION:  
; APPLICANT: MUNN, Edward A.  
; APPLICANT: SMITH, Trevor S.  
; TITLE OF INVENTION: PRODUCTION AND USE OF ANTHELMINTIC  
; AGENTS AND PROTECTIVE IMMUNOGENS  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP  
; STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-5701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/347,198A  
 FILING DATE: 21-NOV-1994  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/020,526  
 FILING DATE: 22-FEB-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/761,749  
 FILING DATE: 17-SEP-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/GB90/00416  
 FILING DATE: 19-MAR-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 89906156.8  
 FILING DATE: 17-MAR-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KITTs, Monica C.  
 REGISTRATION NUMBER: 36,105  
 REFERENCE/DOCKET NUMBER: P443-1289  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202/638-5000  
 TELEFAX: 202/638-4810  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 17 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-347-198A-11

Query Match 0.6%; Score 6; DB 1; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 KAAEVA 886  
 Db 1 KAAEVA 6

RESULT 14  
 US-08-248-839C-53  
 Sequence 53, Application US/08248839C  
 Patent No. 5843702  
 GENERAL INFORMATION:  
 APPLICANT: McConnell, David  
 APPLICANT: Devine, Kevin  
 APPLICANT: O'Kane, Charles  
 TITLE OF INVENTION: A Gene Expression System  
 NUMBER OF SEQUENCES: 185  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 5843702o No. 5843702disk of No. 5843702th America, Inc.  
 STREET: 405 Lexington Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10174-6401  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/248,839C  
 FILING DATE: 25-MAY-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gregg, Valeta A.  
 REGISTRATION NUMBER: 35,127  
 REFERENCE/DOCKET NUMBER: 3614.214-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-867-0123  
 TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 53:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 17 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Protein  
 US-08-248-839C-53

Query Match 0.6%; Score 6; DB 2; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 LTSALI 454  
 Db 5 LTSALI 10

RESULT 15  
 US-08-335-844A-35  
 Sequence 35, Application US/08335844A  
 Patent No. 6066503  
 GENERAL INFORMATION:  
 APPLICANT: GRAHAM, MARGARET  
 APPLICANT: SMITH, TREVOR STANLEY  
 APPLICANT: MUNN, EDWARD ALBERT  
 APPLICANT: KNOX, DAVID PATRICK  
 APPLICANT: OLIVER, JOANNA JANE  
 APPLICANT: NEWTON, SUSAN ELIZABETH  
 TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING  
 TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE  
 NUMBER OF SEQUENCES: 73  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
 STREET: Suite 701-E, 555 Thirteenth St., N.W  
 CITY: Washington  
 STATE: D. C.  
 COUNTRY: U.S.A.  
 ZIP: 20004

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/335,844A  
 FILING DATE: 09-JAN-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB PCT/GB93/00943  
 FILING DATE: 06-MAY-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9209936  
 FILING DATE: 08-MAY-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: WALKER, Barbara W.  
 REGISTRATION NUMBER: 35,400  
 REFERENCE/DOCKET NUMBER: 1181-223A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)783-6040  
 TELEFAX: (202)783-6031  
 INFORMATION FOR SEQ ID NO: 35:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 17 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-335-844A-35

Query Match 0.6%; Score 6; DB 3; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 881 KAAEVA 886

Db 1 KAAEVA 6

Search completed: October 12, 2005, 10:26:31  
Job time : 27 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:18:15 ; Search time 68 Seconds  
(without alignments)  
8072.778 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 1072  
Sequence: 1 MAPPTGVLSLLLVTIAGC.....VSMNGSRNGASFYCSKDR 1072

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 37082

Minimum DB seq length: 0

Maximum DB seq length: 35

Post-processing: Listing first 45 summaries

Database : Uniprot 03:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6	0.6	18	2	Q71UR5	Q71ur5 homo sapien
2	6	0.6	22	2	Q924C7	Q924c7 mus musculus
3	6	0.6	24	2	Q8HWR5	Q8hwr5 homo sapien
4	6	0.6	24	2	Q8HWR6	Q8hwr6 homo sapien
5	6	0.6	24	2	Q8HWR7	Q8hwr7 homo sapien
6	6	0.6	24	2	Q8HWR9	Q8hwr9 homo sapien
7	6	0.6	24	2	Q8HWS0	Q8hws0 homo sapien
8	6	0.6	24	2	Q8HWS1	Q8hws1 homo sapien
9	6	0.6	24	2	Q8HWS2	Q8hws2 homo sapien
10	6	0.6	24	2	Q8MGZ0	Q8mgz0 homo sapien
11	6	0.6	24	2	Q861D7	Q861d7 homo sapien
12	6	0.6	26	1	HCY3_HOMAM	P82298 homarus ame
13	6	0.6	27	1	DMS4_PHYSA	P80280 phyllomedus
14	6	0.6	28	2	Q86618	Q86618 human herpe
15	6	0.6	31	2	Q71VQ3	Q71vq3 oryctolagus
16	6	0.6	32	2	Q7M4E6	Q7m4e6 drosophila
17	6	0.6	33	2	Q52315	Q52315 escherichia
18	6	0.6	35	1	HCYA_CHEDE	P83173 cherax dest
19	6	0.6	35	2	Q8IWM8	Q8iwm8 homo sapien
20	6	0.6	35	2	Q80595	Q80595 human immun
21	5	0.5	9	1	PPH1_LYCES	P83380 lycopersico
22	5	0.5	9	2	Q6UIQ3	Q6uig3 macaca mula
23	5	0.5	9	2	Q70Y80	Q70y80 plectranthu
24	5	0.5	9	2	Q6Q7F3	Q6q7p3 alcaligenes
25	5	0.5	10	2	Q7S225	Q7s225 neurospora
26	5	0.5	10	2	Q68SN8	Q68sm8 chaetodipus
27	5	0.5	10	2	Q68SM9	Q68sm9 thomomys mo
28	5	0.5	10	2	Q68SN0	Q68sn0 cratogeomys
29	5	0.5	10	2	Q68SN1	Q68sn1 cratogeomys
30	5	0.5	10	2	Q6LDC1	Q6ldc1 mus sp. pla
31	5	0.5	10	2	Q39949	Q39949 gb virus c/

32 5 0.5 10 2 Q39957 Q39957 gb virus c/  
33 5 0.5 10 2 Q39958 Q39958 gb virus c/  
34 5 0.5 10 2 Q77VW7 Q77VW7 gb virus c/  
35 5 0.5 10 2 Q77VW8 Q77VW8 gb virus c/  
36 5 0.5 10 2 Q77ZJ1 Q77ZJ1 gb virus c/  
37 5 0.5 10 2 Q77ZJ2 Q77ZJ2 gb virus c/  
38 5 0.5 10 2 Q77ZJ3 Q77ZJ3 gb virus c/  
39 5 0.5 10 2 Q77ZJ4 Q77ZJ4 gb virus c/  
40 5 0.5 10 2 Q9W910 Q9W910 gb virus c/  
41 5 0.5 10 2 Q71HJ9 Q71HJ9 icterus spu  
42 5 0.5 11 2 Q47569 Q47569 escherichia  
43 5 0.5 12 2 Q9X6Y0 Q9X6Y0 aquifex pyr  
44 5 0.5 13 2 Q8I8F2 Q8I8F2 drosophila  
45 5 0.5 13 2 Q7M355 Q7M355 ovis aries

#### ALIGNMENTS

##### RESULT 1

Q71UR5 PRELIMINARY; PRT; 18 AA.  
AC Q71UR5;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Norepinephrine transporter (Fragment).  
GN Name=NET;  
OS Homo sapiens (Human).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99156934; PubMed=10037744; DOI=10.1074/jbc.274.10.6507;  
RA Kim C.H., Kim H.S., Cubellis J.F., Kim K.S.;  
RT "A previously undescribed intron and extensive 5' upstream sequence,  
RT but not Phox2a-mediated transactivation, are necessary for high level  
RT cell type-specific expression of the human norepinephrine transporter  
RT gene.";  
RL J. Biol. Chem. 274:6507-6518 (1999).  
DR EMBL; AF061198; AAD17521.1; -.  
FT NON YER 18 18  
SQ SEQUENCE 18 AA; 1998 MW; E8333908ED35CD2D CRC64;

Query Match 0.6%; Score 6; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 617 VQPENN 622  
Db 10 VQPENN 15  
|||||

##### RESULT 2

Q924C7 PRELIMINARY; PRT; 22 AA.  
AC Q924C7;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Glucagon-like peptide-2 receptor (Fragment).  
GN Name=Glp2r;  
OS Mus musculus (Mouse).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=129/SVJ;  
RC MEDLINE=21292988; PubMed=11262390; DOI=10.1074/jbc.M009382200;  
RA Lovshin J.A., Estall J., Yusta B., Brown T.J., Drucker D.J.;  
RT "Glucagon-like peptide (GLP)-2 action in the murine central nervous

RT system is enhanced by elimination of GLP-1 receptor signaling.";

RL J. Biol. Chem. 276:21489-21499(2001).

DR EMBL; AF338224; AAK63043.1; --.

DR MGD; MGI:2136733; G1p2r.

DR GO; GO:0016021; C:integral to membrane; TAS.

DR GO; GO:0004967; F:glucagon receptor activity; TAS.

KW GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.

FT NON\_TER 22 22

SQ SEQUENCE 22 AA; 2526 MW; 2C5BF53DCCD425C9 CRC64;

Query Match 0.6%; Score 6; DB 2; Length 22;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 SLLLLV 15

Db 13 SLLLLV 18

RESULT 3

Q8HWR5 PRELIMINARY; PRT; 24 AA.

AC Q8HWR5;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DE Leucocyte antigen (Fragment).

GN Name=HLA-B;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ316292; CAC87137.1; --.

FT NON\_TER 24 24

SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332

Db 4 TAPRTV 9

RESULT 4

Q8HWR6 PRELIMINARY; PRT; 24 AA.

AC Q8HWR6;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DE Leucocyte antigen (Fragment).

GN Name=HLA-B;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ316291; CAC87136.1; --.

FT NON\_TER 24 24

SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332

Db 4 TAPRTV 9

RESULT 5

Q8HWR7 PRELIMINARY; PRT; 24 AA.

AC Q8HWR7;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DE Leucocyte antigen (Fragment).

GN Name=HLA-B;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ316290; CAC87135.1; --.

FT NON\_TER 24 24

SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332

Db 4 TAPRTV 9

RESULT 6

Q8HWR9 PRELIMINARY; PRT; 24 AA.

AC Q8HWR9;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DE Leucocyte antigen (Fragment).

GN Name=HLA-B;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ316288; CAC87133.1; --.

FT NON\_TER 24 24

SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332

Db 4 TAPRTV 9

RESULT 7

Q8HWS0 PRELIMINARY; PRT; 24 AA.

AC Q8HWS0;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)



```
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Leucocyte antigen (Fragment).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ316287; CAC87132.1; -.
FT NON_TER 24
SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332
Db 4 TAPRTV 9

RESULT 8
Q8HWS1 PRELIMINARY; PRT; 24 AA.
AC Q8HWS1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Leucocyte antigen (Fragment).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ316286; CAC87879.1; -.
FT NON_TER 24
SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332
Db 4 TAPRTV 9

RESULT 9
Q8HWS2 PRELIMINARY; PRT; 24 AA.
AC Q8HWS2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Leucocyte antigen (Fragment).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ316285; CAC87878.1; -.
FT NON_TER 24
```

```
SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332
Db 4 TAPRTV 9

RESULT 10
Q8MGZ0 PRELIMINARY; PRT; 24 AA.
AC Q8MGZ0;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE Leucocyte antigen (Fragment).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ429190; CAD22131.1; -.
FT NON_TER 24
SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332
Db 4 TAPRTV 9

RESULT 11
Q861D7 PRELIMINARY; PRT; 24 AA.
AC Q861D7;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE MHC class I antigen (Fragment).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ417674; CAD10406.1; -.
FT NON_TER 24
SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332
Db 4 TAPRTV 9

RESULT 12
Q861D7 PRELIMINARY; PRT; 24 AA.
AC Q861D7;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE MHC class I antigen (Fragment).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Swelsen W.T., Voorter C.E., van den Berg-Loonen E.M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ417674; CAD10406.1; -.
FT NON_TER 24
SQ SEQUENCE 24 AA; 2669 MW; 1942CFC2489C177F CRC64;

Query Match 0.6%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 TAPRTV 332
Db 4 TAPRTV 9
```

```

HCY3 HOMAM
ID HCY3 HOMAM STANDARD; PRT; 26 AA.
AC P82298;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemocyanin subunit 3 (Fragment).
OS Homarus americanus (American lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropoidea; Nephropidae; Homarus.
NC NCBI_TaxID=6706;
RN [1]
SEQUENCE.
RP MEDLINE=99259579; PubMed=10327595; DOI=10.1016/S0305-0491(98)10144-X;
RX Stoeva S., Dolashka P., Hristova R., Genov N., Voelter W.;
RA "Subunit composition and N-terminal analysis of arthropod
RT hemocyanins.";
RL Comp. Biochem. Physiol. 122B:69-75(1999).
CC -!- FUNCTION: Hemocyanins are copper-containing oxygen carriers
CC occurring freely dissolved in the hemolymph of many mollusks and
CC arthropods.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Hemolymph.
CC -!- SIMILARITY: Belongs to the tyrosinase family. Hemocyanin
CC subfamily.
DR InterPro: IPR000896; Hemocyanin.
DR InterPro: IPR002227; Tyrosinase.
DR PROSITE; PS00209; HEMOCYANIN 1; PARTIAL.
DR PROSITE; PS00210; HEMOCYANIN 2; PARTIAL.
DR PROSITE; PS00497; TYROSINASE 1; PARTIAL.
DR PROSITE; PS00498; TYROSINASE 2; PARTIAL.
KW Copper; Direct protein sequencing; Hemolymph; Oxygen transport;
KW Transport.
FT NON_TER 26 26
SQ SEQUENCE 26 AA; 2903 MW; 7F3549875A638D61 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 FLLFKV 905
DB 19 FLLFKV 24

RESULT 13
DMS4_PHYSA
ID DMS4_PHYSA STANDARD; PRT; 27 AA.
AC P80280;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Dermaseptin 4 (DS IV).
OS Phyllomedusa sauvagei (Savage's leaf frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae;
OC Phyllomedusinae; Phyllomedusa.
OC NCBI_TaxID=8395;
RN [1]
SEQUENCE.
RP TISSUE=Skin secretion;
RX MEDLINE=94139686; PubMed=8306981;
RA Mor A., Nicolas P.;
RT "Isolation and structure of novel defensive peptides from frog skin.";
RL Eur. J. Biochem. 219:145-154(1994).
CC -!- FUNCTION: Possesses a potent antimicrobial activity against
CC bacteria, fungi and protozoa. Probably acts by disturbing membrane
CC functions with its amphipatic structure.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Skin.
CC -!- SIMILARITY: Belongs to the frog skin active peptide (FSAP) family.
CC Dermaseptin subfamily.

KW Amphibian defense peptide; Antibiotic; Direct protein sequencing;
KW Fungicide; Multigene family.
SQ SEQUENCE 27 AA; 2779 MW; 43C94D2DC19721A8 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 878 KVLKAA 883
DB 9 KVLKAA 14

RESULT 14
Q86618
ID Q86618 PRELIMINARY; PRT; 28 AA.
AC Q86618;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE UL15 protein (Fragment).
GN Name=UL15;
OS Human herpesvirus 1 (HHV-1) (Human herpes simplex virus 1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OC NCBI_TaxID=10298;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=92024477; PubMed=1656627; DOI=10.1016/0168-1702(91)90064-3;
RA Dolan A., Aruckle M., McGeoch D.J.;
RT "Sequence analysis of the splice junction in the transcript of herpes
RT simplex virus type 1 gene UL15.";
RL Virus Res. 20:97-104(1991).
DR EMBL; S62788; AAB20055.1; -.
DR GO; GO:0006323; F:DNA packaging; IEA.
DR InterPro; IPR003499; DNA_pack_N.
DR Pfam; PF02500; DNA_pack_N; 1.
FT NON_TER 28 28
SQ SEQUENCE 28 AA; 3051 MW; CAE2EFDD17528458 CRC64;

Query Match 0.6%; Score 6; DB 2; Length 28;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 DGRSST 744
DB 2 DGRSST 7

RESULT 15
Q71VQ3
ID Q71VQ3 PRELIMINARY; PRT; 31 AA.
AC Q71VQ3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Beta-actin (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand;
RA Carter E.P., Umenishi F., Matthay M.A., Verkman A.S.;
RT "Increased water permeability across the blood-gas barrier in rabbit
RT lungs in the first 24 hours after birth.";
RL J. Clin. Invest. 0:0-0(1997).
DR EMBL; AF000313; AAB94410.1; -.
FT NON_TER 31 31
SQ SEQUENCE 31 AA; 3525 MW; F80D65C92838FA92 CRC64;

```

Query Match 0.6%; Score 6; DB 2; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 DIRKDL 141  
|||  
Db 24 DIRKDL 29

Search completed: October 12, 2005, 10:25:21  
Job time : 69 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:19:24 ; Search time 21 Seconds  
(without alignments)  
4911.637 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 1072  
Sequence: 1 MAPPTGVLSLLLLVVIAGC.....VSMNGSIRNGASFYCSKDR 1072

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 7756

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Listing first 45 summaries

Database : PIR 79:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	0.6	32	2 S55937	translation initia
2	6	0.6	33	2 I53186	IS1/cat protein -
3	6	0.6	35	2 G60529	hemocyanin M3' - c
4	5	0.5	8	2 A21440	variant surface gl
5	5	0.5	10	2 A47384	placental lactogen
6	5	0.5	12	2 P70255	Ig heavy chain CRD
7	5	0.5	12	2 P70319	Ig heavy chain CRD
8	5	0.5	13	2 A38929	glutathione peroxi
9	5	0.5	13	2 S09395	hypothetical prote
10	5	0.5	13	2 S47377	T-cell antigen rec
11	5	0.5	14	2 S29878	Na+/K+-exchanging
12	5	0.5	14	2 S27140	hypothetical prote
13	5	0.5	15	2 S26791	Ig heavy chain V r
14	5	0.5	15	2 P60382	Ig heavy chain J r
15	5	0.5	15	2 G41289	T-cell receptor al
16	5	0.5	15	2 P28587	T-cell receptor be
17	5	0.5	16	2 P50383	Ig heavy chain J r
18	5	0.5	16	2 D58501	26K kidney and gal
19	5	0.5	17	2 C37520	glutathione transf
20	5	0.5	17	2 S26747	Ig heavy chain J r
21	5	0.5	17	2 S26744	Ig heavy chain J r
22	5	0.5	17	2 I24687	T-cell receptor be
23	5	0.5	17	2 P50384	Ig heavy chain J r
24	5	0.5	17	2 A49237	45/47K antigen - M
25	5	0.5	18	2 S24780	protein-tyrosine k
26	5	0.5	18	2 A25941	Ig heavy chain J-H
27	5	0.5	18	4 I39461	anti-angiotensin,
28	5	0.5	19	2 B38837	T-cell receptor be
29	5	0.5	19	2 S28396	T-complex protein

30	5	0.5	19	2	PT0332	Ig heavy chain CRD
31	5	0.5	20	2	A60822	cytochrome P450 PB
32	5	0.5	20	2	S04961	malate dehydrogena
33	5	0.5	20	2	PC1151	equinatoxin 1C - s
34	5	0.5	20	2	B44920	2-halobenzoate 1,2
35	5	0.5	20	2	B44835	dTDPglucose 4,6-de
36	5	0.5	20	2	S63483	coenzyme F420 hydr
37	5	0.5	20	2	B34016	tenebrosin B - sea
38	5	0.5	20	2	B46174	RNA-binding protei
39	5	0.5	21	2	S69371	duodenase - bovine
40	5	0.5	21	2	A61487	110K ATPase, intes
41	5	0.5	21	2	S29858	ribosomal protein
42	5	0.5	21	2	I54268	alpha-1-antichymot
43	5	0.5	22	2	PH1333	Ig heavy chain DJ
44	5	0.5	22	2	D47256	kinetoplast DNA-as
45	5	0.5	23	2	A60529	hemocyanin 1 - edi

ALIGNMENTS

RESULT 1

S55937 translation initiation factor eIF-4E - fruit fly (Drosophila melanogaster) (fragments)  
C:Species: Drosophila melanogaster  
C>Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S55937  
R:Hernandez, G.; Sierra, J.M.  
A:Title: Biochim. Biophys. Acta 1261, 427-431, 1995  
A:Title: Translation initiation factor eIF-4E from Drosophila: cDNA sequence and express  
A:Reference number: S55936; MUID:95260867; PMID:7742371  
A:Accession: S55937  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-8;9-27;28-32 <HER>  
A:Cross-references: UNIPROT:Q7M4E6  
C:Genetics:  
A:Gene: FlyBase:EIF4F  
A:Cross-references: FlyBase:FBgn0013947  
C:Superfamily: translation initiation factor eIF-4E

Query Match 0.6%; Score 6; DB 2; Length 32;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 6; Conservative 0;

Qy	307 PPTSAA 312
Db	21 PPTSAA 26

RESULT 2

I53186 IS1/cat protein - Escherichia coli (fragment)  
C:Species: Escherichia coli  
C>Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 09-Jul-2004  
C:Accession: I53186; I66859  
R:Lida, S.; Marcoli, R.; Bickler, T.A.  
EMBO J. 1, 755-759, 1982  
A:Title: Phenotypic reversion of an IS1-mediated deletion mutation: A combined role for  
A:Reference number: I53186; MUID:84236045; PMID:6329702  
A:Accession: I53186  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-33 <RES>  
A:Cross-references: UNIPROT:Q52315; GB:M24180; NID:g151758; PIDN:AAA26059.1; PID:g151759  
A:Accession: I66859  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-18,'G' <RES2>  
A:Cross-references: GB:M24181; NID:g151760; PIDN:AAA26060.1; PID:g151761

Query Match 0.6%; Score 6; DB 2; Length 33;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 867 IVFYVQ 872  
|||||  
Db 15 IVFYVQ 20

RESULT 3  
G60529  
hemocyanin M3' - crayfish (Cherax destructor) (fragment)  
C:Species: Cherax destructor (Yabby)  
C>Date: 19-Mar-1993 #sequence\_revision 19-Mar-1993 #text\_change 09-Jul-2004  
C:Accession: G60529  
R:Neuteboom, B.; Sierdsema, S.J.; Baintema, J.J.  
Comp. Biochem. Physiol. B 94, 587-592, 1989  
A>Title: The relationship between N-terminal sequences and immunological characterization  
A:Reference number: A60529; MUID:90151075; PMID:2620501  
A:Accession: G60529  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-35 <NEU>  
A:Cross-references: UNIPROT:P83173

Query Match 0.6%; Score 6; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 FLLFKV 905  
|||||  
Db 17 FLLFKV 22

RESULT 4  
A21440  
variant surface glycoprotein pSLc1 - Trypanosoma brucei (fragment)  
C:Species: Trypanosoma brucei  
C>Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jul-2004  
C:Accession: A21440  
R:Parsons, M.; Nelson, R.G.; Watkins, K.P.; Agabian, N.  
Cell 38, 309-316, 1984  
A>Title: Trypanosome mRNAs share a common 5' spliced leader sequence.  
A:Reference number: A90853; MUID:84282716; PMID:6088073  
A:Accession: A21440  
A:Molecule type: mRNA  
A:Residues: 1-8 <PAR>  
A:Cross-references: UNIPROT:P22225; GB:K02195; NID:gl62150; PID:gl62151  
C:Keywords: glycoprotein

Query Match 0.5%; Score 5; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 SGKEV 268  
|||||  
Db 2 SGKEV 6

RESULT 5  
A47364  
placental lactogen-I precursor - mouse (fragment)  
C:Species: Mus sp. (mouse)  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 12-Aug-1996  
C:Accession: A47364  
R:Shida, M.M.; Ng, Y.K.; Soares, M.J.; Linzer, D.I.  
Mol. Endocrinol. 7, 181-188, 1993  
A>Title: Trophoblast-specific transcription from the mouse placental lactogen-I gene pro  
A:Reference number: A47364; MUID:93225959; PMID:8469232  
A:Accession: A47364  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-10 <RES>  
A:Cross-references: GB:S58124; NID:Q299449

Query Match 0.5%; Score 5; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 TLNLS 396  
|||||  
Db 4 TLNLS 8

RESULT 6  
PT0255  
Ig heavy chain CRD3 region (clone 2-115B) - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C:Accession: PT0255  
R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.  
J. Exp. Med. 173, 395-407, 1991  
A>Title: Preferential utilization of specific immunoglobulin heavy chain diversity and  
A:Reference number: PT0222; MUID:91108337; PMID:1899102  
A:Accession: PT0255  
A:Molecule type: DNA  
A:Residues: 1-12 <YAM>  
A:Experimental source: B lymphocyte  
C:Keywords: heterotetramer; immunoglobulin

Query Match 0.5%; Score 5; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 RVSHY 51  
|||||  
Db 3 RVSHY 7

RESULT 7  
PT0319  
Ig heavy chain CRD3 region (clone 6-127) - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C:Accession: PT0319  
R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.  
J. Exp. Med. 173, 395-407, 1991  
A>Title: Preferential utilization of specific immunoglobulin heavy chain diversity and  
A:Reference number: PT0222; MUID:91108337; PMID:1899102  
A:Accession: PT0319  
A:Molecule type: DNA  
A:Residues: 1-12 <YAM>  
A:Experimental source: B lymphocyte  
C:Keywords: heterotetramer; immunoglobulin

Query Match 0.5%; Score 5; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 AVAGP 631  
|||||  
Db 6 AVAGP 10

RESULT 8  
A38929  
glutathione peroxidase (EC 1.11.1.9) - sheep (fragment)  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C>Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 09-Jul-2004  
C:Accession: A38929  
R:Gottlieb, P.; Dyal, D.; Crews, B.  
Arch. Biochem. Biophys. 294, 511-518, 1992  
A>Title: Selenium-dependent glutathione peroxidases from ovine and bovine erythrocytes  
A:Reference number: S21712; MUID:92231574; PMID:1567207  
A:Accession: A38929  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-13 <GET>



A;Cross-references: UNIPROT:Q7M355  
C;Superfamily: glutathione peroxidase  
C;Keywords: oxidoreductase

Query Match 0.5%; Score 5; DB 2; Length 13;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 328 APTV 332  
Db 9 APTV 13

## RESULT 9

S09395  
hypothetical protein - fruit fly (*Drosophila melanogaster*) (fragment)  
C;Species: *Drosophila melanogaster*  
C;Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 24-Jul-1997  
C;Accession: S09395  
R;Gisselmann, G.; Sewing, S.; Madseen, B.W.; Mallart, A.; Angaut-Petit, D.; Mueller-Holtk  
EMBO J. 8, 2359-2364, 1989  
A;Title: The interference of truncated with normal potassium channel subunits leads to a  
A;Reference number: S09395; MUID:90005442; PMID:2551680  
A;Accession: S09395  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-13 <GIS>

Query Match 0.5%; Score 5; DB 2; Length 13;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 SSNS 264  
Db 9 SSNS 13

## RESULT 10

S47377  
T-cell antigen receptor VJ junction beta chain - human  
C;Species: *Homo sapiens* (man)  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 05-Nov-1999  
C;Accession: S47377  
R;Lehner, P.J.  
submitted to the EMBL Data Library, August 1994  
A;Description: Human HLA-A\*0201 restricted recognition of influenza A is dominated by T  
A;Reference number: S47355  
A;Accession: S47377  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-13 <LEH>  
A;Cross-references: EMBL:Z35706; NID:9527505; PIDN:CAA84775.1; PID:9527506  
C;Keywords: T-cell receptor

Query Match 0.5%; Score 5; DB 2; Length 13;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 741 SRSTD 745  
Db 5 SRSTD 9

## RESULT 11

S29878  
Na+/K+-exchanging ATPase (EC 3.6.3.9) alpha-2 chain - rat (fragment)  
C;Species: *Rattus norvegicus* (Norway rat)  
C;Date: 25-Dec-1994 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C;Accession: S29878  
R;Lyttton, J.  
Biochem. Biophys. Res. Commun. 132, 764-769, 1995  
A;Title: The catalytic subunits of the (Na(+), K(+))-ATPase alpha and alpha(+) isozymes  
A;Reference number: S29877; MUID:86050667; PMID:2998384

A;Accession: S29878  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-14 <LVT>  
A;Cross-references: UNIPROT:Q7M0G4  
C;Keywords: hydrolase

Query Match 0.5%; Score 5; DB 2; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 882 AAEVA 886  
Db 7 AAEVA 11

## RESULT 12

S27140  
hypothetical protein 1 estrogen receptor 5'-region - human  
C;Species: *Homo sapiens* (man)  
C;Date: 05-Mar-1994 #sequence\_revision 12-May-1995 #text\_change 16-Aug-2004  
C;Accession: S27140  
R;Keaveney, M.; Klug, J.; Gammon, F.  
DNA Seq. 2, 347-358, 1992  
A;Title: Sequence analysis of the 5' flanking region of the human estrogen receptor gene  
A;Reference number: S27140; MUID:93075998; PMID:1476547  
A;Accession: S27140  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-14 <KEA>  
A;Cross-references: EMBL:X62462; NID:g31201; PIDN:CAA44319.1; PID:g31202

Query Match 0.5%; Score 5; DB 2; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 PSHSL 275  
Db 8 PSHSL 12

## RESULT 13

S26791  
Ig heavy chain V region (N63P2) - human (fragment)  
C;Species: *Homo sapiens* (man)  
C;Date: 13-Jan-1995 #sequence\_revision 02-Aug-1996 #text\_change 20-Jun-2000  
C;Accession: S26791; S19879  
R;Mortari, F.; Newton, J.A.; Wang, J.Y.; Schroeder Jr., H.W.  
Eur. J. Immunol. 22, 241-245, 1992  
A;Title: The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene fami  
A;Reference number: S26786; MUID:92111632; PMID:1730251  
A;Accession: S26791  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-15 <MOR>  
A;Cross-references: EMBL:X61022; NID:g32791; PIDN:CAA43356.1; PID:g1335124; EMBL:X61234;  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotrimer; immunoglobulin

Query Match 0.5%; Score 5; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 407 VTVSS 411  
Db 11 VTVSS 15

## RESULT 14

PS0382  
Ig heavy chain J region 2 - rat (fragment)  
C;Species: *Rattus norvegicus* (Norway rat)  
C;Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 06-Jun-1997

C;Accession: PS0382  
R;Lang, P.; Mocikat, R.  
Gene 102, 261-264, 1991  
A;Title: Immunoglobulin heavy-chain joining genes in the rat: comparison with mouse and  
A;Reference number: JH0666; MUID:91340162; PMID:1908401  
A;Accession: PS0382  
A;Molecule type: DNA  
A;Residues: 1-15 <LAN>  
A;Cross-references: EMBL:X56791  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotrimer; immunoglobulin

Query Match 0.5%; Score 5; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VTVSS 411  
|||  
Db 11 VTVSS 15

## RESULT 15

G41299  
T-cell receptor alpha chain precursor J region (39) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 05-Nov-1999  
C;Accession: G41299  
R;Uematsu, Y.; Wege, H.; Straus, A.; Ott, M.; Bannwarth, W.; Lanchbury, J.; Panayi, G.;  
Proc. Natl. Acad. Sci. U.S.A. 88, 8534-8538, 1991  
A;Title: The T-cell-receptor repertoire in the synovial fluid of a patient with rheumatoid  
A;Reference number: A41299; MUID:92020887; PMID:1656449  
A;Accession: G41299  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-15 <UEM>  
A;Cross-references: GB:S57457; NID:g236330; PIDN:AA19962.1; PID:g236331  
C;Keywords: T-cell receptor

Query Match 0.5%; Score 5; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 692 RLTVK 696  
|||  
Db 10 RLTVK 14

Search completed: October 12, 2005, 10:25:54  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:02:54 ; Search time 73 Seconds  
(without alignments)  
5679.550 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 1072  
Sequence: 1 MAPPTGVLSLLLVTTIAGC.....VSMNGSIRNGASFYCSKDR 1072

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 2105692 seqs, 386760381 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 854512

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Listing first 45 summaries

Database : A Geneseq\_16Dec04.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	1.7	19	4	ABG22066 Novel hum
2	16	1.5	16	8	ADR00632 Human 254
3	14	1.3	24	8	ADR00854 Human 254
4	14	1.3	29	8	ADR00851 Human 254
5	14	1.3	29	8	ADR00857 Human 254
6	9	0.8	9	8	ADR00673 Human 254
7	9	0.8	9	8	ADR00720 Human 254
8	9	0.8	9	8	ADR00722 Human 254
9	9	0.8	9	8	ADR00746 Human 254
10	9	0.8	9	8	ADR00759 Human 254
11	9	0.8	9	8	ADR00767 Human 254
12	9	0.8	9	8	ADR00797 Human 254
13	9	0.8	9	8	ADR00799 Human 254
14	9	0.8	9	8	ADR00835 Human 254
15	9	0.8	9	8	ADR00645 Human 254
16	9	0.8	9	8	ADR00666 Human 254
17	9	0.8	9	8	ADR00683 Human 254
18	9	0.8	9	8	ADR00693 Human 254
19	9	0.8	9	8	ADR00705 Human 254
20	9	0.8	9	8	ADR00706 Human 254
21	9	0.8	9	8	ADR00750 Human 254
22	9	0.8	9	8	ADR00756 Human 254
23	9	0.8	9	8	ADR00772 Human 254
24	9	0.8	9	8	ADR00775 Human 254
25	9	0.8	9	8	ADR00784 Human 254

26	9	0.8	9	8	ADR00790 Human 254
27	9	0.8	9	8	ADR00811 Human 254
28	9	0.8	9	8	ADR00814 Human 254
29	9	0.8	9	8	ADR00844 Human 254
30	9	0.8	9	8	ADR00639 Human 254
31	9	0.8	9	8	ADR00643 Human 254
32	9	0.8	9	8	ADR00648 Human 254
33	9	0.8	9	8	ADR00656 Human 254
34	9	0.8	9	8	ADR00671 Human 254
35	9	0.8	9	8	ADR00672 Human 254
36	9	0.8	9	8	ADR00680 Human 254
37	9	0.8	9	8	ADR00688 Human 254
38	9	0.8	9	8	ADR00689 Human 254
39	9	0.8	9	8	ADR00694 Human 254
40	9	0.8	9	8	ADR00721 Human 254
41	9	0.8	9	8	ADR00741 Human 254
42	9	0.8	9	8	ADR00754 Human 254
43	9	0.8	9	8	ADR00760 Human 254
44	9	0.8	9	8	ADR00782 Human 254
45	9	0.8	9	8	ADR00805 Human 254

ALIGNMENTS

RESULT 1  
ABG22066  
XX ABG22066 standard; protein; 19 AA.  
XX AC ABG22066;  
XX DT 18-FEB-2002 (first entry)  
XX DB Novel human diagnostic protein #22057.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX FN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS86253.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.  
Claim 20; SEQ ID NO 52425; 103pp; English.  
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (III). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 19 AA;

Query Match 1.7%; Score 18; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-09; Mismatches 0; Indels 0; Gaps 0;  
 Matches 18; Conservative 0;

QY 1 MAPPTGVLSSLLLVITIA 18  
 |||||  
 DB 1 MAPPTGVLSSLLLVITIA 18

RESULT 2  
 ADR00632  
 ID ADR00632 standard; peptide; 16 AA.  
 XX  
 AC ADR00632;

XX 04-NOV-2004 (first entry)

XX Human 254PID6B peptide SEQ ID NO:43.

XX 254PID6B; small interfering RNA; siRNA; immune response;  
 KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.

XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX Example 9; SEQ ID NO 43; 345pp; English.

XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID  
 CC NO:3.5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of PI; (2) a polynucleotide

CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of PI; (5)  
 CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
 CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of PI; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of PI; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of PI. 254PID6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254PID6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254PID6B gene is located on chromosome 6p22.  
 XX

SQ Sequence 16 AA;

Query Match 1.5%; Score 16; DB 8; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-07; Mismatches 0; Indels 0; Gaps 0;  
 Matches 16; Conservative 0;

QY 522 NNAVDPVPVANAGPNH 537  
 |||||  
 DB 1 NNAVDPVPVANAGPNH 16

RESULT 3  
 ADR00854

ID ADR00854 standard; peptide; 24 AA.

XX ADR00854;

XX 04-NOV-2004 (first entry)

XX Human 254PID6B v.3 peptide SEQ ID NO:265.

XX 254PID6B; small interfering RNA; siRNA; immune response;  
 KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254PID6B v.3; chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.

XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX Example 13; SEQ ID NO 265; 345pp; English.

XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,

CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (p1, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of p1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of p1; (5)  
 CC detecting, in a sample, the presence of a 254p1D6B-related protein or a  
 CC 254p1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of p1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of p1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of p1. 254p1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents a  
 CC human 254p1D6B v.3 peptide, which is used in the exemplification of the  
 CC present invention. The human 254p1D6B gene is located on chromosome 6p22.  
 CC  
 CC Sequence 24 AA;

Query Match 1.3%; Score 14; DB 8; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 CARQKQCEGRYYSN 33  
 Db 11 CARQKQCEGRYYSN 24  
 |||||

RESULT 4  
 ADR00851  
 ID ADR00851 standard; peptide; 29 AA.

XX ADR00851;  
 XX  
 XX 04-NOV-2004 (first entry)  
 DT Human 254p1D6B v.2 peptide SEQ ID NO:262.  
 DE  
 DE 254p1D6B; small interfering RNA; siRNA; immune response;  
 KW 254p1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254p1D6B v.2; chromosome 6.  
 XX  
 XX Homo sapiens.

OS  
 XX WO2004067716-A2.  
 XX 12-AUG-2004.  
 FD  
 XX 23-JAN-2004; 2004WO-US001965.  
 PF  
 XX 24-JAN-2003; 2003US-0442526P.  
 XX  
 XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;  
 XX WPI; 2004-580991/56.  
 DR  
 XX New 254p1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254p1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX Example 13; SEQ ID NO 262; 345pp; English.  
 PS  
 XX

CC The present invention describes a 254p1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254p1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (p1, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of p1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of p1; (5)  
 CC detecting, in a sample, the presence of a 254p1D6B-related protein or a  
 CC 254p1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of p1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of p1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of p1. 254p1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents a  
 CC human 254p1D6B v.2 peptide, which is used in the exemplification of the  
 CC present invention. The human 254p1D6B gene is located on chromosome 6p22.  
 CC  
 CC Sequence 29 AA;

Query Match 1.3%; Score 14; DB 8; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 FLGKDWGLEMSEY 156  
 Db 1 FLGKDWGLEMSEY 14  
 |||||

RESULT 5  
 ADR00857  
 ID ADR00857 standard; peptide; 29 AA.

XX ADR00857;  
 XX  
 XX 04-NOV-2004 (first entry)  
 DT Human 254p1D6B v.5 peptide SEQ ID NO:268.  
 XX  
 DE 254p1D6B; small interfering RNA; siRNA; immune response;  
 KW 254p1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 KW 254p1D6B v.5; chromosome 6.  
 XX  
 XX Homo sapiens.

OS  
 XX WO2004067716-A2.  
 XX 12-AUG-2004.  
 PD  
 XX 23-JAN-2004; 2004WO-US001965.  
 PF  
 XX 24-JAN-2003; 2003US-0442526P.  
 XX  
 XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
 PI Perez-Villar JJ, Faris M;  
 XX WPI; 2004-580991/56.  
 DR  
 XX New 254p1D6B siRNA composition comprising a double stranded siRNA that

PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.

XX Example 13; SEQ ID NO 268; 345pp; English.

XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of P1; (5)  
CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of P1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of P1. 254PID6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents a  
CC human 254PID6B v.5 peptide, which is used in the exemplification of the  
CC present invention. The human 254PID6B gene is located on chromosome 6p22.

XX Sequence 29 AA;

Query Match 1.3%; Score 14; DB 8; Length 29;

Best Local Similarity 100.0%; Pred. NO. 3.1e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 GIWGDSPEIRKDL 141

Db 1 GIWGDSPEIRKDL 14

RESULT 6

ADR00673

ID ADR00673 standard; peptide; 9 AA.

XX ADR00673;

XX 04-NOV-2004 (first entry)

XX Human 254PID6B peptide SEQ ID NO:84.

XX 254PID6B; small interfering RNA; siRNA; immune response;

KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
KW chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US0001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX

PI Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;  
PI Perez-Villarr JJ, Paris M;  
XX WPI; 2004-580991/56.

XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.

PS Example 9; SEQ ID NO 84; 345pp; English.

XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of P1; (5)  
CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of P1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of P1. 254PID6B has cytostatic activity, and can be  
CC used in gene therapy. The compositions, molecules and methods are useful  
CC for treating and detecting cancer. The present sequence represents the  
CC human 254PID6B peptide, which is used in the exemplification of the  
CC present invention. The human 254PID6B gene is located on chromosome 6p22.

XX Sequence 9 AA;

Query Match 0.8%; Score 9; DB 8; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.8e+06;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 320 ELPISPTTA 328

Db 1 ELPISPTTA 9

RESULT 7

ADR00720

ID ADR00720 standard; peptide; 9 AA.

XX ADR00720;

XX 04-NOV-2004 (first entry)

XX Human 254PID6B peptide SEQ ID NO:131.

XX 254PID6B; small interfering RNA; siRNA; immune response;

KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;  
KW chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX





DE	Human 254PID6B peptide SEQ ID NO:157.
XX	
XX	254PID6B; small interfering RNA; siRNA; immune response;
KW	254PID6B-related protein; cytostatic; gene therapy; cancer; human;
KW	chromosome 6.
XX	
XX	Homo sapiens.
OS	
XX	WO2004067716-A2.
PN	
XX	12-AUG-2004.
PD	
XX	
XX	23-JAN-2004; 2004WO-US001965.
PF	
XX	
XX	24-JAN-2003; 2003US-0442526P.
PR	
XX	(AGEN-) AGENSYS INC.
PA	
XX	
PI	Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;
PI	Perez-Villar JJ, Faris M;
XX	
DR	WPI; 2004-580991/56.
XX	
XX	New 254PID6B siRNA composition comprising a double stranded siRNA that
PT	corresponds to the nucleic acid ORF sequence which encodes the 254PID6B
PT	protein or corresponds to a subsequence of the ORF, useful for detecting
PT	and treating cancer.
XX	
XX	Example 9; SEQ ID NO 157; 345pp; English.
PS	
XX	
CC	The present invention describes a 254PID6B small interfering RNA (siRNA)
CC	composition that comprises a double stranded siRNA that corresponds to
CC	the nucleic acid open reading frame (ORF) sequence which encodes the
CC	254PID6B protein, or corresponds to a subsequence of the ORF, where the
CC	double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous
CC	nucleotides in length. Also described: (1) a composition that comprises,
CC	consists essentially of, or consists of a peptide of eight, nine, ten, or
CC	eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID
CC	NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in
CC	any of the 42 lists of peptides, given in the specification, or a protein
CC	that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or
CC	identical to an entire amino acid sequence of P1; (2) a polynucleotide
CC	that encodes the protein; (3) a composition comprising a polynucleotide
CC	that is a full complement of the polynucleotide described above; (4)
CC	generating a mammalian immune response directed to the protein of P1; (5)
CC	detecting, in a sample, the presence of a 254PID6B-related protein or a
CC	254PID6B-related polynucleotide; (6) a composition that modulates the
CC	status of a cell that expresses a protein of P1; (7) a pharmaceutical
CC	composition that comprises the composition described above in a human
CC	unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)
CC	a non-human transgenic animal that produces the antibody; (10) a
CC	hybridoma that produces the antibody; (11) delivering a cytotoxic agent
CC	or a diagnostic agent to a cell that expresses the protein of P1; and
CC	(12) inhibiting growth, reproduction or survival of cancer cells that
CC	express the protein of P1. 254PID6B has cytostatic activity, and can be
CC	used in gene therapy. The compositions, molecules and methods are useful
CC	for treating and detecting cancer. The present sequence represents the
CC	human 254PID6B peptide, which is used in the exemplification of the
CC	present invention. The human 254PID6B gene is located on chromosome 6p22.
XX	
XX	Sequence 9A:
XX	

ID	ADR00759	standard; peptide; 9 AA.
XX		
AC	ADR00759;	
AC		
XX		
DT	04-NOV-2004	(first entry)
DT		
XX		
XX	Human 254PID6B peptide	SEQ ID NO:170.
DE		
XX		
XX	254PID6B; small interfering RNA; siRNA; immune response;	
KW	254PID6B-related protein; cytostatic; gene therapy; cancer; human;	
KW	chromosome 6.	
KW		
XX		
OS	Homo sapiens.	
XX		
XX	WO2004067716-A2.	
FN		
XX		
PD	12-AUG-2004.	
XX		
XX		
PF	23-JAN-2004; 2004WO-US001965.	
XX		
XX	24-JAN-2003; 2003US-0442526P.	
PR		
XX	(AGEN-) AGENSYS INC.	
PA		
XX		
PI	Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;	
PI	Perez-Villar JJ, Faris M;	
XX		
DR	WPI; 2004-580991/56.	
XX		
PT	New 254PID6B siRNA composition comprising a double stranded siRNA that	
PT	corresponds to the nucleic acid ORF sequence which encodes the 254PID6B	
PT	protein or corresponds to a subsequence of the ORF, useful for detecting	
PT	and treating cancer.	
XX		
PS	Example 9; SEQ ID NO 170; 345pp; English.	
XX		
CC	The present invention describes a 254PID6B small interfering RNA (siRNA)	
CC	composition that comprises a double stranded siRNA that corresponds to	
CC	the nucleic acid open reading frame (ORF) sequence which encodes the	
CC	254PID6B protein, or corresponds to a subsequence of the ORF, where the	
CC	double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous	
CC	nucleotides in length. Also described: (1) a composition that comprises,	
CC	consists essentially of, or consists of a peptide of eight, nine, ten, or	
CC	eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID	
CC	NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in	
CC	any of the 42 lists of peptides, given in the specification, or a protein	
CC	that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or	
CC	identical to an entire amino acid sequence of P1; (2) a polynucleotide	
CC	that encodes the protein; (3) a composition comprising a polynucleotide	
CC	that is a full complement of the polynucleotide described above; (4)	
CC	generating a mammalian immune response directed to the protein of P1; (5)	
CC	detecting, in a sample, the presence of a 254PID6B-related protein or a	
CC	254PID6B-related polynucleotide; (6) a composition that modulates the	
CC	status of a cell that expresses a protein of P1; (7) a pharmaceutical	
CC	composition that comprises the composition described above in a human	
CC	unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)	
CC	generating a mammalian immune response directed to the protein of P1; (10) a	
CC	non-human transgenic animal that produces the antibody; (11) delivering a cytotoxic agent	
CC	hybridoma that produces the antibody; (12) delivering a cytotoxic agent	
CC	or a diagnostic agent to a cell that expresses the protein of P1; and	
CC	(13) inhibiting growth, reproduction or survival of cancer cells that	
CC	express the protein of P1. 254PID6B has cytostatic activity, and can be	
CC	used in gene therapy. The compositions, molecules and methods are useful	
CC	for treating and detecting cancer. The present sequence represents the	
CC	human 254PID6B peptide, which is used in the exemplification of the	
CC	present invention. The human 254PID6B gene is located on chromosome 6p22.	
CC		

Db 1 PTPPTSAAP 9  
|||||  
Query Match 0.8%; Score 9; DB 8; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11  
ID ADR00767 standard; peptide; 9 AA.  
XX ADR00767;  
AC ADR00767;  
XX 04-NOV-2004 (first entry)  
XX Human 254P1D6B peptide SEQ ID NO:178.  
XX 254P1D6B; small interfering RNA; siRNA; immune response;  
KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
KW chromosome 6.  
XX Homo sapiens.  
XX WO2004067716-A2.  
XX 12-AUG-2004.  
XX 23-JAN-2004; 2004WO-US001965.  
XX 24-JAN-2003; 2003US-0442526P.  
XX (AGEN-) AGENSYS INC.  
XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
PI Perez-Villar JJ, Paris M;  
XX WPI; 2004-580991/56.  
XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX Example 9; SEQ ID NO 178; 345pp; English.

XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of P1; (5)  
CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of P1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
XX Sequence 9 AA;

Query Match 0.8%; Score 9; DB 8; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 363 PVETTYNYE 371  
| | | | |  
Db 1 PVETTYNYE 9

RESULT 12  
ID ADR00797 standard; peptide; 9 AA.  
XX ADR00797;  
AC ADR00797;  
XX 04-NOV-2004 (first entry)  
XX Human 254P1D6B peptide SEQ ID NO:208.  
XX 254P1D6B; small interfering RNA; siRNA; immune response;  
KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
KW chromosome 6.  
XX Homo sapiens.  
XX WO2004067716-A2.  
XX 12-AUG-2004.  
XX 23-JAN-2004; 2004WO-US001965.  
XX 24-JAN-2003; 2003US-0442526P.  
XX (AGEN-) AGENSYS INC.  
XX Kanner SB, Raitano AB, Jakobovits A, Challita-Bid PM, Ge W;  
PI Perez-Villar JJ, Paris M;  
XX WPI; 2004-580991/56.  
XX New 254P1D6B siRNA composition comprising a double stranded siRNA that  
PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
PT protein or corresponds to a subsequence of the ORF, useful for detecting  
PT and treating cancer.  
XX Example 9; SEQ ID NO 208; 345pp; English.

XX The present invention describes a 254P1D6B small interfering RNA (siRNA)  
CC composition that comprises a double stranded siRNA that corresponds to  
CC the nucleic acid open reading frame (ORF) sequence which encodes the  
CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
CC nucleotides in length. Also described: (1) a composition that comprises,  
CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
CC any of the 42 lists of peptides, given in the specification, or a protein  
CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
CC that encodes the protein; (3) a composition comprising a polynucleotide  
CC that is a full complement of the polynucleotide described above; (4)  
CC generating a mammalian immune response directed to the protein of P1; (5)  
CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
CC composition that comprises the composition described above in a human  
CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
CC a non-human transgenic animal that produces the antibody; (10) a  
CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
CC or a diagnostic agent to a cell that expresses the protein of P1; and  
CC (12) inhibiting growth, reproduction or survival of cancer cells that  
CC express the protein of P1. 254P1D6B has cytostatic activity, and can be

CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254PID6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254PID6B gene is located on chromosome 6p22.

SQ Sequence 9 AA;

Query Match 0.8%; Score 9; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0;

QY 733 NNSITIDGS 741  
 |||||

Db 1 NNSITIDGS 9

RESULT 13

ADRO0799

ID ADRO0799 standard; peptide; 9 AA.

XX ADR00799;

DT 04-NOV-2004 (first entry)

DE Human 254PID6B peptide SEQ ID NO:210.

XX 254PID6B; small interfering RNA; siRNA; immune response;

KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;

KW chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;

PI Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.

XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX Example 9; SEQ ID NO 210; 345pp; English.

XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of PI; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of PI; (5)  
 CC detecting, in a sample, the presence of a 254PID6B-related protein or a  
 CC 254PID6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of PI; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human

CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of PI; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of PI. 254PID6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254PID6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254PID6B gene is located on chromosome 6p22.

SQ Sequence 9 AA;

Query Match 0.8%; Score 9; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0;

QY 775 ALQLTNLVE 783

Db 1 ALQLTNLVE 9

RESULT 14

ADRO0835

ID ADRO0835 standard; peptide; 9 AA.

XX ADR00835;

DT 04-NOV-2004 (first entry)

DE Human 254PID6B peptide SEQ ID NO:246.

XX 254PID6B; small interfering RNA; siRNA; immune response;

KW 254PID6B-related protein; cytostatic; gene therapy; cancer; human;

KW chromosome 6.

XX Homo sapiens.

XX WO2004067716-A2.

XX 12-AUG-2004.

XX 23-JAN-2004; 2004WO-US001965.

XX 24-JAN-2003; 2003US-0442526P.

XX (AGEN-) AGENSYS INC.

XX Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;

PI Perez-Villar JJ, Faris M;

XX WPI; 2004-580991/56.

XX New 254PID6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254PID6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX Example 9; SEQ ID NO 246; 345pp; English.

XX The present invention describes a 254PID6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254PID6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or  
 CC eleven contiguous amino acids of a protein of figure 2 (PI, see SEQ ID  
 CC NO:3, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of PI; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide

CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254P1D6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 XX  
 SQ Sequence 9 AA;

Query Match 0.8%; Score 9; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 558 QIVLYEWSL 566  
 Db 1 QIVLYEWSL 9  
 |||||

RESULT 15  
 ADRO0645  
 ID ADRO0645 standard; peptide; 9 AA.

XX AC ADRO0645;

XX DT 04-NOV-2004 (first entry)

XX DE Human 254P1D6B peptide SEQ ID NO:56.

XX KW 254P1D6B; small interfering RNA; siRNA; immune response;

XX KW 254P1D6B-related protein; cytostatic; gene therapy; cancer; human;  
 XX KW chromosome 6.

XX OS Homo sapiens.

XX PN WO2004067716-A2.

XX PD 12-AUG-2004.

XX PF 23-JAN-2004; 2004WO-US001965.

XX PR 24-JAN-2003; 2003US-0442526P.

XX PA (AGEN-) AGENSYS INC.

XX PI Kanner SB, Raitano AB, Jakobovits A, Challita-Eid PM, Ge W;

XX PI Perez-Villar JJ, Farris M;

XX DR WPI; 2004-580991/56.

XX PT New 254P1D6B siRNA composition comprising a double stranded siRNA that  
 PT corresponds to the nucleic acid ORF sequence which encodes the 254P1D6B  
 PT protein or corresponds to a subsequence of the ORF, useful for detecting  
 PT and treating cancer.

XX PS Example 9; SEQ ID NO 56; 345pp; English.

XX CC The present invention describes a 254P1D6B small interfering RNA (siRNA)  
 CC composition that comprises a double stranded siRNA that corresponds to  
 CC the nucleic acid open reading frame (ORF) sequence which encodes the  
 CC 254P1D6B protein, or corresponds to a subsequence of the ORF, where the  
 CC double stranded siRNA is 19, 20, 21, 22, 23, 24, or 25 contiguous  
 CC nucleotides in length. Also described: (1) a composition that comprises,  
 CC consists essentially of, or consists of a peptide of eight, nine, ten, or

CC eleven contiguous amino acids of a protein of figure 2 (P1, see SEQ ID  
 CC NO:13, 5 or 7 ADR00592, ADR00594 or ADR00596), or a peptide included in  
 CC any of the 42 lists of peptides, given in the specification, or a protein  
 CC that is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homologous or  
 CC identical to an entire amino acid sequence of P1; (2) a polynucleotide  
 CC that encodes the protein; (3) a composition comprising a polynucleotide  
 CC that is a full complement of the polynucleotide described above; (4)  
 CC generating a mammalian immune response directed to the protein of P1; (5)  
 CC detecting, in a sample, the presence of a 254P1D6B-related protein or a  
 CC 254P1D6B-related polynucleotide; (6) a composition that modulates the  
 CC status of a cell that expresses a protein of P1; (7) a pharmaceutical  
 CC composition that comprises the composition described above in a human  
 CC unit dose form; (8) an antibody or its fragment, which is monoclonal; (9)  
 CC a non-human transgenic animal that produces the antibody; (10) a  
 CC hybridoma that produces the antibody; (11) delivering a cytotoxic agent  
 CC or a diagnostic agent to a cell that expresses the protein of P1; and  
 CC (12) inhibiting growth, reproduction or survival of cancer cells that  
 CC express the protein of P1. 254P1D6B has cytostatic activity, and can be  
 CC used in gene therapy. The compositions, molecules and methods are useful  
 CC for treating and detecting cancer. The present sequence represents the  
 CC human 254P1D6B peptide, which is used in the exemplification of the  
 CC present invention. The human 254P1D6B gene is located on chromosome 6p22.  
 XX  
 SQ Sequence 9 AA;

Query Match 0.8%; Score 9; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 IWGDSPEDI 137  
 Db 1 IWGDSPEDI 9  
 |||||

Search completed: October 12, 2005, 10:24:01  
 Job time : 74 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:27:56 ; Search time 70 Seconds

(without alignments)  
6370.973 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPRTGVLSILLLVITAGC.....VSMNGSIRNGASPSYCSKDR 1072

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1854112 seqs, 416015017 residues

Total number of hits satisfying chosen parameters: 458990

Minimum DB seq length: 0

Maximum DB seq length: 35

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

```

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	158	2.8	29	16 US-10-764-390-268	Sequence 268, App
2	152	2.7	29	16 US-10-764-390-262	Sequence 262, App
3	101	1.8	19	16 US-10-764-390-261	Sequence 261, App
4	99	1.8	19	16 US-10-764-390-267	Sequence 267, App
5	92	1.6	16	16 US-10-764-390-43	Sequence 43, App
6	90	1.6	17	16 US-10-764-390-266	Sequence 266, App
7	84	1.5	17	16 US-10-764-390-260	Sequence 260, App
8	83	1.5	19	18 US-10-450-763-52425	Sequence 52425, A
9	79	1.4	24	16 US-10-764-390-265	Sequence 265, App
10	63	1.1	9	16 US-10-764-390-138	Sequence 138, App
11	61.5	1.1	34	9 US-09-892-835-1	Sequence 1, Appli

12	61.5	1.1	34	10	US-09-769-180-3	Sequence 3, Appli
13	61.5	1.1	34	15	US-10-682-103-3	Sequence 3, Appli
14	60	1.1	12	16	US-10-764-390-41	Sequence 41, Appl
15	59	1.1	9	16	US-10-764-390-137	Sequence 137, App
16	59	1.1	9	16	US-10-764-390-249	Sequence 249, App
17	57	1.0	9	16	US-10-764-390-52	Sequence 52, Appl
18	57	1.0	31	10	US-09-984-130-64	Sequence 64, Appl
19	57	1.0	31	10	US-09-836-353A-64	Sequence 64, Appl
20	57	1.0	34	14	US-10-029-386-30913	Sequence 30913, A
21	56	1.0	9	16	US-10-764-390-139	Sequence 139, App
22	56	1.0	9	16	US-10-764-390-225	Sequence 225, App
23	56	1.0	9	16	US-10-764-390-240	Sequence 240, App
24	55	1.0	9	16	US-10-764-390-50	Sequence 50, Appl
25	55	1.0	9	16	US-10-764-390-86	Sequence 86, Appl
26	55	1.0	9	16	US-10-764-390-239	Sequence 239, App
27	55	1.0	11	16	US-10-764-390-42	Sequence 42, Appl
28	55	1.0	27	15	US-10-038-854-390	Sequence 390, App
29	55	1.0	32	14	US-10-062-831-177	Sequence 177, App
30	55	1.0	32	14	US-10-062-599-177	Sequence 177, App
31	55	1.0	33	10	US-09-984-130-58	Sequence 58, Appl
32	55	1.0	33	10	US-09-836-353A-58	Sequence 58, Appl
33	54	1.0	9	16	US-10-764-390-105	Sequence 105, App
34	54	1.0	9	16	US-10-764-390-136	Sequence 136, App
35	54	1.0	9	16	US-10-764-390-158	Sequence 158, App
36	54	1.0	9	16	US-10-764-390-221	Sequence 221, App
37	54	1.0	9	16	US-10-764-390-237	Sequence 237, App
38	54	1.0	29	15	US-10-176-419A-4	Sequence 4, Appli
39	54	1.0	34	10	US-09-984-130-62	Sequence 62, Appl
40	54	1.0	34	10	US-09-836-353A-62	Sequence 62, Appl
41	53.5	1.0	24	14	US-10-292-896-12	Sequence 12, Appl
42	53	0.9	9	16	US-10-764-390-55	Sequence 55, Appl
43	53	0.9	9	16	US-10-764-390-56	Sequence 56, Appl
44	53	0.9	9	16	US-10-764-390-95	Sequence 95, Appl
45	53	0.9	9	16	US-10-764-390-177	Sequence 177, App

#### ALIGNMENTS

#### RESULT 1

```

US-10-764-390-268
; Sequence 268, Application US/10764390
; Publication No. US2004021421A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Rid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Faris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 268
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-268

```

Query Match 2.8%; Score 158; DB 16; Length 29;  
Best Local Similarity 96.6%; Pred. NO. 0.0062;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 128 GINGDSPEDIRKDLXFLGKQWGLEMSY 156

```
Db 1 GIGDSEDIRKDLTFLGKDWGLEMSY 29
|||||
RESULT 2
US-10-764-390-262
; Sequence 262, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Bid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 262
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-262

Query Match 2.7%; Score 152; DB 16; Length 29;
Best Local Similarity 96.6%; Pred. No. 0.015;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 143 FLGKDWGLEMSYDDYRELEKDLQPS 171
|||||
Db 1 FLGKDWGLEMSYADDYRELEKDLQPS 29
|||||

RESULT 3
US-10-764-390-261
; Sequence 261, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Bid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-261

Query Match 1.8%; Score 101; DB 16; Length 19;
Best Local Similarity 94.7%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

Qy 148 WGLEMSYDDYRELEKD 166
|||||
Db 1 WGLEMSYADDYRELEKD 19
|||||

Matches 18; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

Qy 148 WGLEMSYDDYRELEKD 166
|||||
Db 1 WGLEMSYADDYRELEKD 19
|||||

RESULT 4
US-10-764-390-267
; Sequence 267, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Bid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 267
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-267

Query Match 1.8%; Score 99; DB 16; Length 19;
Best Local Similarity 94.7%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 133 SPEDIRKDLTFLGKDWGLE 151
|||||
Db 1 SPEDIRKDLTFLGKDWGLE 19
|||||

RESULT 5
US-10-764-390-43
; Sequence 43, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Bid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-43
```



Query Match 1.6%; Score 92; DB 16; Length 16;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 522 NNAVDYPPVANAGPNH 537  
Db 1 NNAVDYPPVANAGPNH 16

## RESULT 6

US-10-764-390-266  
; Sequence 266, Application US/10764390  
; Publication No. US20040214212A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Jakobovits, Aya  
; APPLICANT: Challita-Eid, Pia M.  
; APPLICANT: Ge, Wangmao  
; APPLICANT: Faris, Mary  
; APPLICANT: Steven B. Kanner  
; APPLICANT: Juan J. Perez-Villar  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20081.00  
; CURRENT APPLICATION NUMBER: US/10/764,390  
; CURRENT FILING DATE: 2004-01-23  
; PRIOR APPLICATION NUMBER: US60/442,526  
; PRIOR FILING DATE: 2003-01-24  
; NUMBER OF SEQ ID NOS: 277  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 266  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-764-390-266

Query Match 1.6%; Score 90; DB 16; Length 17;  
Best Local Similarity 94.1%; Pred. No. 45;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 134 PEDIRKDLXFLGKDWGL 150  
Db 1 PEDIRKDLTFLGKDWGL 17

## RESULT 7

US-10-764-390-260  
; Sequence 260, Application US/10764390  
; Publication No. US20040214212A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Jakobovits, Aya  
; APPLICANT: Challita-Eid, Pia M.  
; APPLICANT: Ge, Wangmao  
; APPLICANT: Faris, Mary  
; APPLICANT: Steven B. Kanner  
; APPLICANT: Juan J. Perez-Villar  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20081.00  
; CURRENT APPLICATION NUMBER: US/10/764,390  
; CURRENT FILING DATE: 2004-01-23  
; PRIOR APPLICATION NUMBER: US60/442,526  
; PRIOR FILING DATE: 2003-01-24  
; NUMBER OF SEQ ID NOS: 277  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 260  
; LENGTH: 17

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-764-390-260

Query Match 1.5%; Score 84; DB 16; Length 17;  
Best Local Similarity 94.1%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 149 GLEEMSEYDDYRELEK 165  
Db 1 GLEEMSEYDDYRELEK 17

## RESULT 8

US-10-450-763-52425  
; Sequence 52425, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 03/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 52425  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(19)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-52425

Query Match 1.5%; Score 83; DB 18; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPPTGVLSLLLVTTIA 18  
Db 1 MAPPTGVLSLLLVTTIA 18

## RESULT 9

US-10-764-390-265  
; Sequence 265, Application US/10764390  
; Publication No. US20040214212A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Jakobovits, Aya  
; APPLICANT: Challita-Eid, Pia M.  
; APPLICANT: Ge, Wangmao  
; APPLICANT: Faris, Mary  
; APPLICANT: Steven B. Kanner  
; APPLICANT: Juan J. Perez-Villar  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 254PID6B Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20081.00  
; CURRENT APPLICATION NUMBER: US/10/764,390  
; CURRENT FILING DATE: 2004-01-23  
; PRIOR APPLICATION NUMBER: US60/442,526  
; PRIOR FILING DATE: 2003-01-24  
; NUMBER OF SEQ ID NOS: 277  
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 265
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-265

Query Match      1.4%; Score 79; DB 16; Length 24;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CARQCSGRTYSN 33
    |||||
DB 11 CARQCSGRTYSN 24
    |||||

RESULT 10
US-10-764-390-138
; Sequence 138, Application US/10764390
; Publication No. US20040214212A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Ge, Wangmao
; APPLICANT: Paris, Mary
; APPLICANT: Steven B. Kanner
; APPLICANT: Juan J. Perez-Villar
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 254Plb6B Useful in Treatment and Detection of
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 51158-20081.00
; CURRENT APPLICATION NUMBER: US/10/764,390
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US60/442,526
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-390-138

Query Match      1.1%; Score 63; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 946 WDGESNCEW 954
    |||||
DB 1 WDGESNCEW 9

RESULT 11
US-09-892-835-1
; Sequence 1, Application US/09892835
; Publication No. US2002019016A1
; GENERAL INFORMATION:
; APPLICANT: Hulstaert, Frank
; APPLICANT: Vanmechelen, Eugene
; APPLICANT: Vanderstichele, Hugo
; TITLE OF INVENTION: Differential Diagnosis of Neurological Diseases
; FILE REFERENCE: INNS:027 (11362.0027.NPUS01)
; CURRENT APPLICATION NUMBER: US/09/892,835
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 09/892.835
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/218.907
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: EP 00870151.8
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 1
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: MOD RES
; LOCATION: (28)..(28)
; OTHER INFORMATION: PHOSPHORYLATION THREONINE
US-09-892-835-1

Query Match      1.1%; Score 61.5; DB 9; Length 34;
Best Local Similarity 29.6%; Pred. No. 6.8e+03;
Matches 21; Conservative 6; Mismatches 7; Indels 37; Gaps 4;

QY 176 PRGSAEYTDWGLLPQSEGAFFNSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVL 235
    |||||
DB 1 PRGAAP-----PQKQGANAT-----RIPAKT-----PPAPKTP----- 29
    |||||

QY 236 LPLPTTPSSGE 246
    |||||
DB 30 -----PSSGE 34

RESULT 12
US-09-769-180-3
; Sequence 3, Application US/09769180
; Publication No. US20030194742A1
; GENERAL INFORMATION:
; APPLICANT: VanMechelen, Eugene
; APPLICANT: Vanderstichele, Hugo
; TITLE OF INVENTION: Diagnosis of Tauopathies
; FILE REFERENCE: US.112.T181
; CURRENT APPLICATION NUMBER: US/09/769,180
; CURRENT FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: EP 00870008.8
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: EP 00870280.5
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/178,391
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (28)..(28)
; OTHER INFORMATION: PHOSPHORYLATION
US-09-769-180-3

Query Match      1.1%; Score 61.5; DB 10; Length 34;
Best Local Similarity 29.6%; Pred. No. 6.8e+03;
Matches 21; Conservative 6; Mismatches 7; Indels 37; Gaps 4;

QY 176 PRGSAEYTDWGLLPQSEGAFFNSVGDSPAVPAETQDDPELHYLNESASTPAPKLPERSVL 235
    |||||
DB 1 PRGAAP-----PQKQGANAT-----RIPAKT-----PPAPKTP----- 29
    |||||

QY 236 LPLPTTPSSGE 246
    |||||
DB 30 -----PSSGE 34

RESULT 13
US-10-682-103-3
; Sequence 3, Application US/10682103
; Publication No. US20040091942A1
; GENERAL INFORMATION:
; APPLICANT: VanMechelen, Eugene
```

Qy 151 EEMSEYXDDYRE 162

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:26:35 ; Search time 26 Seconds  
(without alignments)  
3077.838 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 5580  
Sequence: 1 MAPPTGVLSLLLVITAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 250370

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61.5	1.1	34	US-09-892-835-1	Sequence 1, Appli
2	61.5	1.1	34	US-09-769-180-3	Sequence 3, Appli
3	59.5	1.1	35	US-08-658-136-52	Sequence 52, Appli
4	55	1.0	32	US-09-690-454-177	Sequence 177, App
5	54	1.0	25	US-09-060-767B-4	Sequence 4, Appli
6	53.5	1.0	24	US-09-217-306B-19	Sequence 19, Appli
7	51	0.9	31	US-08-602-999A-54	Sequence 54, Appli
8	51	0.9	31	US-08-278-865-54	Sequence 54, Appli
9	51	0.9	31	US-09-500-124-54	Sequence 54, Appli
10	51	0.9	31	US-09-938-315-54	Sequence 54, Appli
11	49	0.9	22	US-09-430-564-9	Sequence 9, Appli
12	49	0.9	22	US-09-430-564-15	Sequence 15, Appli
13	49	0.9	23	US-08-833-807-6	Sequence 6, Appli
14	49	0.9	23	US-09-223-043-6	Sequence 6, Appli
15	49	0.9	23	US-09-593-870A-6	Sequence 6, Appli
16	49	0.9	30	US-09-039-780A-87	Sequence 87, Appli
17	49	0.9	31	US-08-244-951A-4	Sequence 4, Appli
18	49	0.9	31	US-09-280-047-10	Sequence 10, Appli
19	49	0.9	31	US-08-208-573B-10	Sequence 10, Appli
20	49	0.9	33	US-08-389-011-2	Sequence 2, Appli
21	49	0.9	33	US-08-403-917A-2	Sequence 2, Appli
22	49	0.9	33	US-09-348-952A-2	Sequence 10, Appli
23	49	0.9	34	US-08-602-264A-10	Sequence 10, Appli
24	49	0.9	34	US-08-461-018A-10	Sequence 10, Appli
25	49	0.9	34	US-09-216-958-10	Sequence 10, Appli
26	49	0.9	35	US-08-244-951A-6	Sequence 6, Appli
27	48.5	0.9	18	US-09-910-009A-401	Sequence 401, App

28	48	0.9	17	4	US-09-910-009A-395	Sequence 395, App
29	48	0.9	17	4	US-09-910-009A-413	Sequence 413, App
30	48	0.9	17	4	US-09-910-009A-447	Sequence 447, App
31	48	0.9	18	4	US-09-910-009A-451	Sequence 451, App
32	48	0.9	21	4	US-09-646-154-1	Sequence 1, Appli
33	48	0.9	27	2	US-08-394-600B-4	Sequence 4, Appli
34	48	0.9	27	4	US-08-385-456C-4	Sequence 4, Appli
35	48	0.9	27	4	US-08-487-453A-4	Sequence 4, Appli
36	48	0.9	27	5	PCT-US95-02513-4	Sequence 9, Appli
37	48	0.9	30	3	US-09-050-811-9	Sequence 9, Appli
38	48	0.9	34	2	US-08-942-423-16	Sequence 16, Appli
39	47	0.8	27	4	US-09-636-399A-17	Sequence 17, Appli
40	47	0.8	29	4	US-09-636-399A-15	Sequence 15, Appli
41	47	0.8	30	4	US-09-636-399A-16	Sequence 16, Appli
42	47	0.8	30	4	US-09-636-399A-72	Sequence 72, Appli
43	47	0.8	30	4	US-09-270-767-59044	Sequence 59044, A
44	47	0.8	30	4	US-09-270-767-52474	Sequence 62474, A
45	47	0.8	31	4	US-09-636-399A-70	Sequence 70, Appli

ALIGNMENTS

RESULT 1  
US-09-892-835-1  
; Sequence 1, Application US/09892835  
; Patent No. 6670137  
; GENERAL INFORMATION:  
; APPLICANT: Hulstaert, Frank  
; APPLICANT: Vanmechelen, Eugene  
; APPLICANT: Vanderstichelen, Hugo  
; TITLE OF INVENTION: Differential Diagnosis of Neurological Diseases  
; FILE REFERENCE: INNS:027 (11362.0027.NPUS01)  
; CURRENT APPLICATION NUMBER: US/09/892.835  
; CURRENT FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: US 09/892.835  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: US 60/218.907  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: EP 00870151.8  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 34  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Synthetic Peptide  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (28)..(28)  
; OTHER INFORMATION: PHOSPHORYLATION THREONINE  
US-09-892-835-1

Query Match	1.1%;	Score	61.5;	DB	4;	Length	34;
Best Local Similarity	29.6%;	Pred. No.	4.1e+02;				
Matches	21;	Conservative	6;	Mismatches	7;	Indels	37;
Gaps	4;						
Qy	176	PRGSAETDGLLPGSGAFNSVGDSPVAPAEATQDDPDLHYLINESASTPAPKLPERSVL	235				
Db	1	PRGAAP	1				
Qy	236	LPLPTPSSGE	246				
Db	30	-----PSSGE	34				

RESULT 2  
US-09-769-180-3  
; Sequence 3, Application US/09769180  
; Patent No. 6680173  
; GENERAL INFORMATION:

```

; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-658-136-52

Query Match 1.1%; Score 59.5; DB 3; Length 35;
Best Local Similarity 35.0%; Pred. No. 6.2e+02;
Matches 14; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY 560 VLYEWSLGFSGSGKHVVMQGVQTPYHLHSAMQEGDYTFQL 599
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 VAYHWFDDGSGSPG-----QDTDEPRAEHSYLRFGDYRVQV 35

RESULT 4
US-09-690-454-177
; Sequence 177, Application US/09690454
; Patent No. 6531447
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: P2006P1
; CURRENT APPLICATION NUMBER: US/09/690,454
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: 09/189,144
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/056,250
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,296
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,293
; PRIOR FILING DATE: August 29, 1997
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 177
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-690-454-177

Query Match 1.0%; Score 55; DB 4; Length 32;
Best Local Similarity 45.2%; Pred. No. 1.2e+03;
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 155 EYXDDYRELEKOLLQSPGKQEPGRGSABYTDW 185
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 EYNDYTELGLQKLKESGKQ--HGFAFSFDY 29

RESULT 5
US-09-060-767B-4
; Sequence 4, Application US/09060767B
; Patent No. 6720152
; GENERAL INFORMATION:
; APPLICANT: Well, Gary
; APPLICANT: Chandrashekar, Ramaswamy
; TITLE OF INVENTION: Diagnosis of Histoplasmosis Using Antigens Specific for

```

;; TITLE OF INVENTION: H. capsulatum  
;; FILE REFERENCE: BJCH 9986  
;; CURRENT APPLICATION NUMBER: US/09/060,767B  
;; CURRENT FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/043,332  
;; PRIOR FILING DATE: 1997-04-15  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 4  
;; LENGTH: 25  
;; TYPE: PRT  
;; ORGANISM: Caldocellum saccharolyticum  
US-09-060-767B-4

Query Match 1.0%; Score 54; DB 4; Length 25;  
Best Local Similarity 45.8%; Pred. No. 9.8e+02;  
Matches 11; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 291 PVLTVPGSTHSIPTPTSAAPS 314  
Db 1 PTSTVTPTPTPTPTPTVTATPT 24

RESULT 6  
US-09-217-306B-19  
; Sequence 19, Application US/09217306B  
; Patent No. 6465220  
; GENERAL INFORMATION:  
; APPLICANT: Hassan, Helle  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Bennett, Eric P.  
; TITLE OF INVENTION: Glycosylation Using GalNAc-T4 Transferase  
; FILE REFERENCE: 8850\*1  
; CURRENT APPLICATION NUMBER: US/09/217,306B  
; CURRENT FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: Muc7  
US-09-217-306B-19

Query Match 1.0%; Score 53.5; DB 4; Length 24;  
Best Local Similarity 50.0%; Pred. No. 1e+03;  
Matches 11; Conservative 4; Mismatches 4; Indels 3; Gaps 1;

Qy 296 TPGSTHSIPTPTSAAPSEST 317  
Db 4 TPSAT---TPAPPSSAPPET 22

RESULT 7  
US-08-602-999A-54  
; Sequence 54, Application US/08602999A  
; Patent No. 6184205  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: KAY, Brian K.  
; APPLICANT: THORN, Judith M.  
; APPLICANT: QUILLIAM, Lawrence A.  
; APPLICANT: DER, Channing J.  
; APPLICANT: FOWLKES, Dana M.  
; APPLICANT: RIDER, James B.  
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
; TITLE OF INVENTION: ISOLATING AND USING SAME  
; NUMBER OF SEQUENCES: 467  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas

;; CITY: New York  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/602,999A  
;; FILING DATE: 16-FEB-1996  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Misrock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 1101-202  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 54:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 31 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: peptide  
US-08-602-999A-54

Query Match 0.9%; Score 51; DB 3; Length 31;  
Best Local Similarity 62.5%; Pred. No. 2.4e+03;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 314 SESTPSELPISTPTAP 329  
Db 14 SRSTPRPLPLPTTP 29

RESULT 8  
US-08-278-865-54  
; Sequence 54, Application US/08278865  
; Patent No. 6303574  
; GENERAL INFORMATION:  
; APPLICANT: KAY, BRIAN K.  
; APPLICANT: SPARKS, ANDREW B.  
; APPLICANT: THORN, JUDITH M.  
; APPLICANT: QUILLIAM, LAWRENCE A.  
; APPLICANT: DER, CHANNING J.  
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
; TITLE OF INVENTION: ISOLATING AND USING SAME  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/278,865  
;; FILING DATE:  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Villacorta, Gilberto M.  
;; REGISTRATION NUMBER: 34,038  
;; REFERENCE/DOCKET NUMBER: 4980-007-0  
;; TELECOMMUNICATION INFORMATION:



TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEFAX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-278-865-54

Query Match 0.9%; Score 51; DB 3; Length 31;  
Best Local Similarity 62.5%; Pred. No. 2.4e+03;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 314 SESTPSELPISPTTAP 329  
DB 14 SRSTPRPLMLPTTRP 29

RESULT 9  
US-09-500-124-54  
Sequence 54, Application US/09500124  
Patent No. 6432920  
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: QUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
NUMBER OF SEQUENCES: 467  
TITLE OF INVENTION: ISOLATING AND USING SAME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/500,124  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,999  
FILING DATE: 16-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mierock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-09-500-124-54

Query Match 0.9%; Score 51; DB 4; Length 31;  
Best Local Similarity 62.5%; Pred. No. 2.4e+03;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 314 SESTPSELPISPTTAP 329  
DB 14 SRSTPRPLMLPTTRP 29

RESULT 10  
US-09-938-315-54  
Sequence 54, Application US/09938315  
Patent No. 6703482  
GENERAL INFORMATION:

APPLICANT: KAY, BRIAN K.  
SPARKS, ANDREW B.  
THORN, JUDITH M.  
QUILLIAM, LAWRENCE A.  
DER, CHANNING J.

TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/938,315

FILING DATE: 23-Aug-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Villacorta, Gilberto M.

REGISTRATION NUMBER: 34,038

REFERENCE/DOCKET NUMBER: 4980-007-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: Peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 54:

US-09-938-315-54

Query Match 0.9%; Score 51; DB 4; Length 31;  
Best Local Similarity 62.5%; Pred. No. 2.4e+03;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 314 SESTPSELPISPTTAP 329  
DB 14 SRSTPRPLMLPTTRP 29

RESULT 11  
US-09-430-564-9  
Sequence 9, Application US/09430564  
Patent No. 6372467  
GENERAL INFORMATION:

APPLICANT: John Blenis

APPLICANT: Kay K. Lee-Fruman

APPLICANT: Calvin J. Kuo

TITLE OF INVENTION: P54S6K AND P85S6K GENES, PROTEINS,

TITLE OF INVENTION: PRIMERS, PROBES, AND DETECTION METHODS

FILE REFERENCE: 00246/506002

```
; CURRENT APPLICATION NUMBER: US/09/430,564
; CURRENT FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/106,141
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-430-564-9

Query Match      0.9%  Score 49;  DB 3;  Length 22;
Best Local Similarity 47.6%;  Pred. No. 2e+03;
Matches 10;  Conservative 1;  Mismatches 10;  Indels 0;  Gaps 0;

Qy 305 PTPPTSAAPSESTPSELPISP 325
Db 1 PLPPLPPPPPTTAPLPFIRP 21

RESULT 12
US-09-430-564-15
; Sequence 15, Application US/09430564
; Patent No. 6372467
; GENERAL INFORMATION:
; APPLICANT: John Blenis
; APPLICANT: Kay K. Lee-Fruman
; APPLICANT: Calvin J. Kuo
; TITLE OF INVENTION: P54S6K AND P85S6K GENES, PROTEINS,
; FILE REFERENCE: 00246/506002
; CURRENT APPLICATION NUMBER: US/09/430,564
; CURRENT FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/106,141
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-430-564-15

Query Match      0.9%  Score 49;  DB 3;  Length 22;
Best Local Similarity 47.6%;  Pred. No. 2e+03;
Matches 10;  Conservative 1;  Mismatches 10;  Indels 0;  Gaps 0;

Qy 305 PTPPTSAAPSESTPSELPISP 325
Db 1 PLPPLPPPPPTTAPLPFIRP 21

RESULT 13
US-08-833-807-6
; Sequence 6, Application US/08833807
; Patent No. 5989552
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/223,043
; APPLICATION NUMBER: US/09/223,043
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/833,807
; FILING DATE:
; APPLICATION NUMBER: AU PM3223
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagen, Patrick J.
; REGISTRATION NUMBER: 27,643
; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,807
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/833,807
; FILING DATE:
; APPLICATION NUMBER: AU PM3223
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagen, Patrick J.
; REGISTRATION NUMBER: 27,643
; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR

; Query Match      0.9%  Score 49;  DB 2;  Length 23;
; Best Local Similarity 40.9%;  Pred. No. 2.2e+03;
; Matches 9;  Conservative 4;  Mismatches 9;  Indels 0;  Gaps 0;

Qy 294 TVTPGSTSHSIPTPTSAAPSE 315
Db 2 TAAPTPTTATPPPPSSAPPE 23

RESULT 14
US-09-223-043-6
; Sequence 6, Application US/09223043
; Patent No. 6177256
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/223,043
; APPLICATION NUMBER: US/09/223,043
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/833,807
; FILING DATE:
; APPLICATION NUMBER: AU PM3223
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagen, Patrick J.
; REGISTRATION NUMBER: 27,643
; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215)563-4100  
TELEFAX: (215)563-4044  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-223-043-6

Query Match 0.9%; Score 49; DB 3; Length 23;  
Best Local Similarity 40.9%; Pred. No. 2.2e+03;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 294 TVTPGSTSHSIPTPTSAAPSE 315  
Db 2 TAAPPTPPATTPAPSSAPPE 23

RESULT 15  
US-09-593-870A-6  
Sequence 6, Application US/09593870A  
Patent No. 6548643  
GENERAL INFORMATION:  
APPLICANT: McKenzie, Ian F. C.  
APPLICANT: Apostolopoulos, Vasso  
APPLICANT: Pietersz, Geoff Allan  
TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their  
FILE REFERENCE: 2368-McKenzie  
CURRENT APPLICATION NUMBER: US/09/593,870A  
CURRENT FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: 09/223,043  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-593-870A-6

Query Match 0.9%; Score 49; DB 4; Length 23;  
Best Local Similarity 40.9%; Pred. No. 2.2e+03;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 294 TVTPGSTSHSIPTPTSAAPSE 315  
Db 2 TAAPPTPPATTPAPSSAPPE 23

Search completed: October 12, 2005, 10:31:46  
Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:25:25 ; Search time 67 Seconds

(without alignments)  
8193.267 Million cell updates/sec

Title: US-10-764-390-3

Perfect score: 5580

Sequence: 1 MAPPTGVLSLLLVTTIAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 37082

Minimum DB seq length: 0

Maximum DB seq length: 35

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	51	0.9	32	Q7M4E6	Q7M4E6 drosophila
2	49.5	0.9	34	Q6PJK3	Q6PJK3 homo sapien
3	48	0.9	27	Q8CLU0	Q8CLU0 yersinia pe
4	48	0.9	35	TXH5 ORNHU	P61104 ornithocton
5	47.5	0.9	27	TXA3 ANESU	P01535 anemonia su
6	47	0.8	18	Q6LID4	Q6LID4 homo sapien
7	47	0.8	21	Q8WLR1	Q8WLR1 homo sapien
8	47	0.8	23	SKL1_TITCA	P83243 titiys camb
9	46	0.8	24	Q7YP58	Q7YP58 homo sapien
10	46	0.8	29	Q6WY52	Q6WY52 bacterioph
11	46	0.8	32	TAT_SIVM2	P05912 simian immu
12	46	0.8	34	Q70222	Q70222 human immu
13	45.5	0.8	35	Q9YMB1	Q9YMB1 human immu
14	45	0.8	27	Q9Y181	Q9Y181 priapulid c
15	45	0.8	29	Q7R7E1	Q7R7E1 plasmidius
16	45	0.8	33	Q87JTB9	Q87JTB9 vibrio para
17	45	0.8	33	Q39816	Q39816 encephalomy
18	44.5	0.8	34	Q735G0	Q735G0 bacillus ce
19	44.5	0.8	35	Q80586	Q80586 human immu
20	44	0.8	34	Q8J3X2	Q8J3X2 human immu
21	44	0.8	35	SCXP ANDMA	P01498 androctonus
22	43.5	0.8	20	Q8IXG1	Q8IXG1 homo sapien
23	43.5	0.8	33	Q8EFK7	Q8EFK7 shewanella
24	43.5	0.8	35	Q74EN3	Q74EN3 geobacter s
25	43	0.8	22	Q85607	Q85607 streptomyce
26	43	0.8	26	Q38675	Q38675 bacterioph
27	43	0.8	27	Q77D54	Q77D54 bacterioph
28	43	0.8	27	Q9AZ16	Q9AZ16 bacterioph
29	43	0.8	34	Q7Y043	Q7Y043 vitis vinif
30	43	0.8	34	Q2ZG81	Q2ZG81 chlamydia t
31	43	0.8	35	Q83333	Q83333 murine hepa

32	42.5	0.8	32	1	RK1_RABIT	P81655 oryctolagus
33	42.5	0.8	34	2	Q7RK45	Q7RK45 plasmodium
34	42.5	0.8	35	2	Q7RX56	Q7RX56 neurospora
35	42.5	0.8	35	2	Q9Y083	Q9Y083 globodera r
36	42.5	0.8	35	2	Q77900	Q77900 human immu
37	42.5	0.8	35	2	Q80555	Q80555 human immu
38	42.5	0.8	35	2	Q80562	Q80562 human immu
39	42.5	0.8	35	2	Q80563	Q80563 human immu
40	42.5	0.8	35	2	Q80564	Q80564 human immu
41	42.5	0.8	35	2	Q80565	Q80565 human immu
42	42	0.8	22	2	Q7M3F1	Q7M3F1 bos taurus
43	42	0.8	23	2	Q8HA12	Q8HA12 bacterioph
44	42	0.8	24	1	LPER_STRFR	P45440 streptomyce
45	42	0.8	25	2	Q7Y286	Q7Y286 phage phi 4

#### ALIGNMENTS

##### RESULT 1

ID	Q7M4E6	PRELIMINARY;	PRT;	32 AA.
AC	Q7M4E6;			
DT	01-MAR-2004 (TrEMBLrel. 26, Created)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Translation initiation factor eIF-4E (Fragments).			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE.			
RA	Hernandez G., Sierra J.M.;			
RT	"Translation initiation factor eIF-4E from Drosophila: CDNA sequence			
RT	and expression of the gene."			
RL	Biochim. Biophys. Acta 1261:427-431 (1995).			
DR	PIR, S55937; S55937.			
FT	NON TER 32			
SQ	SEQUENCE 32 AA; 3364 MW; C492C82B551A6959 CRC64;			

Query Match 0.9%; Score 51; DB 2; Length 32;  
Best Local Similarity 68.8%; Pred. No. 3e+04;  
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 299 STEHSIPTPTSAAPS 314

Db 13 STEQGRPEPTSAAS 28

##### RESULT 2

ID	Q6PJK3	PRELIMINARY;	PRT;	34 AA.
AC	Q6PJK3;			
DT	05-JUL-2004 (TrEMBLrel. 27, Created)			
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)			
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)			
DE	PM1G protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.2426038999;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			

```

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smailus D.E., Schermer A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014386; AAH14386.2; -
SQ SEQUENCE 34 AA; 3886 MW; 7B38FEFC8B7809D4 CRC64;

Query Match 0.9%; Score 48.5; DB 2; Length 34;
Best Local Similarity 37.0%; Pred. No. 4.4e+04;
Matches 10; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

QY 915 LKCSGHGCHDPLTKRCICSHLWMENLI 941
Db 1 MKQSGFGYCHPLWKSCWIS-AWHQTL 26

RESULT 3
Q8CLU0 PRELIMINARY; PRT; 27 AA.
ID Q8CLU0
AC Q8CLU0
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical.
GN OrderedLocNames=y0241;
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RX DOI=10.1128/JB.184.16.4601-4611.2002;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Milles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AE013624; AAM83835.1; -
KW Hypothetical protein.
SQ SEQUENCE 27 AA; 3211 MW; A5203621B3DE0ABC CRC64;

Query Match 0.9%; Score 48; DB 2; Length 27;
Best Local Similarity 39.1%; Pred. No. 3.4e+04;
Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 841 LAVLLNVLDSDIKVKQIRASDL 863
Db 4 LLLIYQRTDSEIKLYRVGSHSL 26

RESULT 4
TXHS_ORNHU
ID TXHS_ORNHU
AC P61104;

Query Match 0.9%; Score 48; DB 1; Length 35;
Best Local Similarity 31.2%; Pred. No. 4.8e+04;
Matches 10; Conservative 3; Mismatches 11; Indels 8; Gaps 1;

```

```

DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Huwentoxin-V (HwTx-V) [Contains: Mutant of huwentoxin-V (mHWTX-V)].
OS Ornithotoxus huwena (Chinese bird spider) (Selenocosmia huwena).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Ornithoctonus.
OX NCBI_TaxID=29017;
RN [1]
RP SEQUENCE, DISULFIDE BONDS, AND MASS SPECTROMETRY.
RC TISSUE=Venom;
RX PubMed=12893056; DOI=10.1016/S0041-0101(03)00095-3;
RA Zhang P.-F., Chen P., Hu W.-J., Liang S.-P.;
RT "Huwentoxin-V, a novel insecticidal peptide toxin from the spider
RT Selenocosmia huwena, and a natural mutant of the toxin: indicates the
RT key amino acid residues related to the biological activity.";
RL Toxicon 42:15-20(2003).
CC -!- FUNCTION: HwTx-V reversibly paralyzes locusts and cockroaches, and
CC causes death with larger doses. It has no effect on mice by intra-
CC abdominal, nor intracerebroventricular injection. Its natural
CC mutant mHwTx-V shows no effect on locusts, cockroaches, and mice.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -!- MASS SPECTROMETRY: MW=4111.4; MW_ERR=0.4; METHOD=MALDI; RANGE=1-
CC 35; NOTE=Ref.1.
CC -!- MASS SPECTROMETRY: MW=3877.1; MW_ERR=0.4; METHOD=MALDI; RANGE=1-
CC 33; NOTE=Ref.1.
CC -!- MISCELLANEOUS: ED(50) of HwTx-V is 16 +/- 5 mg/kg to locusts.
KW Direct protein sequencing; Neurotoxin; Toxin.
FT CHAIN 1 35 Huwentoxin-V.
FT CHAIN 1 33 Mutant of huwentoxin-V.
FT DISULFID 2 16 By similarity.
FT DISULFID 9 21 By similarity.
FT DISULFID 15 28 By similarity.
SQ SEQUENCE 35 AA; 4117 MW; CC39F421571CA523 CRC64;

Query Match 0.9%; Score 48; DB 1; Length 35;
Best Local Similarity 31.2%; Pred. No. 4.8e+04;
Matches 10; Conservative 3; Mismatches 11; Indels 8; Gaps 1;

QY 917 CSGHGCHDPLTKRCICSHLWMENLIQRYWDG 948
Db 9 CSQGDGC-----CKHLQCHSNYWCWVDG 32

RESULT 5
TXA3_ANESU
ID TXA3_ANESU STANDARD; PRT; 27 AA.
AC P01535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Neurotoxin III (Toxin ATX-III).
OS Anemonia sulcata (Snake-locks sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaeae; Actiniidae; Anemonia.
OX NCBI_TaxID=6108;
RN [1]
RP SEQUENCE.
RC TISSUE=Cnidoblast;
RX MEDLINE=78084776; PubMed=23311; DOI=10.1016/0014-5793(77)80699-6;
RA Martinez G., Kopeyan C., Schweitz H., Lazdunski M.;
RT "Toxin III from Anemonia sulcata: primary structure.";
RL FEBS Lett. 84:247-252(1977).
RN [2]
RP PRELIMINARY SEQUENCE.
RX MEDLINE=78044787; PubMed=21843;
RA Beress L., Wunderer G., Wachter E.;
RT "Amino acid sequence of toxin III from Anemonia sulcata.";
RL Hoppe-Seyler's Z. Physiol. Chem. 358:985-988(1977).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=93343891; PubMed=8102051;

```

RA Norton R.S., Cross K., Braach-Maksvytis V., Wachter E.;  
RT "1H-NMR study of the solution properties and secondary structure of  
RL neurotoxin III from the sea anemone Anemonia sulcata.";  
RN Biochem. J. 293:545-551(1993).  
[4]  
RN STRUCTURE BY NMR.  
RX MEDLINE=95244415; PubMed=7727358;  
RA Manoleras N., Norton R.S.;  
RT "Three-dimensional structure in solution of neurotoxin III from the  
RT sea anemone Anemonia sulcata.";  
RL Biochemistry 33:11051-11061(1994).  
CC -!- FUNCTION: Binds specifically to the sodium channel.  
CC -!- SUBCELLULAR LOCATION: Secreted; cnicocyst.  
CC -!- SIMILARITY: Belongs to the sea anemone short toxin family.  
DR PIR; A91446; TZA23.  
DR PDB; 1ANS; NMR; @=1-27.  
KW 3D-structure; Cnidocyst; Direct protein sequencing;  
KW Ionic channel inhibitor; Neurotoxin; Sodium channel inhibitor; Toxin.  
FT DISULFID 3 17  
FT DISULFID 4 11  
FT DISULFID 6 22  
FT CONFLICT 22 23 CS -> SC (in Ref. 2).  
FT TURN 7 10  
SQ SEQUENCE 27 AA; 2938 MW; AA4E261FFAF34A7A CRC64;  
Query Match 0.9%; Score 47.5; DB 1; Length 27;  
Best Local Similarity 28.1%; Pred. No. 3.6e+04;  
Matches 9; Conservative 3; Mismatches 7; Indels 13; Gaps 1;  
QY 60 ACCDLSDDLAWFGRGCVLYSCPHKENCEPK 91  
DB 2 SCCP-----CYWGGCPWQNCYPE 20  
RESULT 6  
QSLD14  
ID Q6LD14 PRELIMINARY; PRT; 18 AA.  
AC Q6LD14;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Tyrosine kinase (Fragment).  
GN Name=TXK;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Peripheral blood;  
RX MEDLINE=95038742; PubMed=7951233;  
RA Haire R.N., Ohta Y., Lewis J.E., Fu S.M., Kroisel P., Litman G.W.;  
RT "TXK, a novel human tyrosine kinase expressed in T cells shares  
RL sequence identity with Tec family kinases and maps to 4p12.";  
RN Hum. Mol. Genet. 3:897-901(1994).  
[2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Peripheral blood;  
RA Litman G.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U07791; AAA19597.1; -.  
DR GO; GO:0016301; F:kinase activity; IEA.  
KW Kinase.  
FT NON\_TER 1 1  
FT NON\_TER 18 18  
SQ SEQUENCE 18 AA; 2025 MW; D954E7BB497625B1 CRC64;  
Query Match 0.8%; Score 47; DB 2; Length 18;  
Best Local Similarity 70.0%; Pred. No. 2.2e+04;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 977 CICCCKRQKR 986  
| | | | |

Db 9 CCCCOSVQKR 18  
RESULT 7  
Q8WLR1  
ID Q8WLR1 PRELIMINARY; PRT; 21 AA.  
AC Q8WLR1;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Major histocompatibility complex class I (Fragment).  
GN Name=HLA-B;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=97246738; PubMed=9089103; DOI=10.1007/s002510050227;  
RA Vargas-Alarcon G., Gomez-Casado B., Martinez-Laso J., Granados J.,  
RA Layrisse Z., Alegre R., Arnaiz-Villena A.;  
RT "Differences in intron 2 sequences between B\*39061 and B\*39062 in  
RT Amerindians: comparison with those of B\*3901, B\*5101, and B\*52012  
RT alleles".  
RL Immunogenetics 45:436-439(1997).  
DR EMBL; I77205; AAL40076.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 21 21  
SQ SEQUENCE 21 AA; 1972 MW; 5941EC53ECF3A73D CRC64;  
Query Match 0.8%; Score 47; DB 2; Length 21;  
Best Local Similarity 56.2%; Pred. No. 2.7e+04;  
Matches 9; Conservative 4; Mismatches 1; Indels 2; Gaps 1;  
QY 622 NRPPVAV--AGPKDEL 635  
| | | | | : : : : :  
DB 4 NRPPAALGGSGPDRDL 19  
RESULT 8  
SCK1\_TITCA  
ID SCK1\_TITCA STANDARD; PRT; 23 AA.  
AC P83243;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Toxin Tc1 (Alpha-KTx 13.1).  
OS Tityus cambridgei (Amazonian scorpion).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
OC Buthida; Buthoidea; Buthidae; Tityus.  
OX NCBI\_TaxID=184226;  
[1]  
RN SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.  
RC TISSUE=Venom;  
RX MEDLINE=20565534; PubMed=1113450; DOI=10.1016/S0014-5793(00)02253-5;  
RA Batista C.V.F., Gomez-Lagunas F., Lucas S., Possani L.D.;  
RT "Tc1, from Tityus cambridgei, is the first member of a new subfamily  
RT of scorpion toxin that blocks K(+) channels.";  
RL FEBS Lett. 486:117-120(2000).  
[2]  
RN STRUCTURE BY NMR.  
RX PubMed=11790849; DOI=10.1110/ps.33402;  
RA Wang I., Wu S.H., Chang H.K., Shieh R.C., Yu H.M., Chen C.;  
RT "Solution structure of a K(+) channel blocker from the scorpion Tityus  
RT cambridgei.";  
RL Protein Sci. 11:390-400(2002).  
CC -!- FUNCTION: Blocks reversibly Shaker B potassium channels. Also  
CC displaces binding of noxiustoxin to mouse brain synaptosome  
CC membranes.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
CC -!- MASS SPECTROMETRY: MW=2446.4; METHOD=MALDI; RANGE=1-23;

```

RX MEDLINE=22990536; PubMed=14629033;
RA Bull J.J., Badgett M.R., Rokytá D., Molineux I.J.;
RT "Experimental evolution yields hundreds of mutations in a functional
RL viral genome.";
RL J. Mol. Evol. 57:241-248(2003).
DR ENBL; AY264778; AAP34129.1; -.
DR ENBL; AY264777; AAP34075.1; -.
SQ SEQUENCE 29 AA; 3217 MW; 5EF0F978F980B9D1 CRC64;

Query Match 0.8%; Score 46; DB 2; Length 29;
Best Local Similarity 47.1%; Pred. No. 4.7e+04;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 11 LLLLVTIAGCARKQCSE 27
:||||: |||
Db 5 MLLLIIVGCLALYCSD 21

RESULT 11
TAT_SIVM2 STANDARD; PRT; 32 AA.
ID AC P05912;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE TAT protein (Transactivating regulatory protein) (Fragment) .
OS Name=TAT;
OS Simian immunodeficiency virus (Mm251 isolate) (SIV-MAC) .
OS Viruses; Retrovird viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11734;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=68122665; PubMed=2893293; DOI=10.1038/331619a0;
RA Kestler H.W., Li Y., Naidu Y.M., Butler C.V., Ochs M.F., Jaenel G.,
RA King N.W., Daniel M.D., Desrosiers R.C.;
RT "Comparison of simian immunodeficiency virus isolates.";
RL Nature 331:619-622(1988) .
CC -I- FUNCTION: Transcriptional regulator that acts by binding to the
CC trans-activating responsive sequence (TAR) RNA element and
CC activates transcription initiation and/or elongation from the LTR
CC promoter.
CC CC -I- SUBUNIT: Binds cyclin T1 (By similarity) .
CC CC -I- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC CC -I- MISCELLANEOUS: This is a macaque isolate.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC ENBL; X06879; -; NOT ANNOTATED_CDS.
CC DR HIV; X06879; TAT$MM251.
CC DR Activator; AIDS; Nuclear protein; RNA-binding;
CC KW Transcription regulation.
CC FT NON_TER 1 1
SQ SEQUENCE 32 AA; 3556 MW; 211547CBE6F4D817 CRC64;

Query Match 0.8%; Score 46; DB 1; Length 32;
Best Local Similarity 42.1%; Pred. No. 5.3e+04;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 666 PSAVEMENIDKAIATVTGL 684
| : | : | : | : |
Db 12 PERAKKETVEKAVATAPGL 30

RESULT 12
Q70222 PRELIMINARY; PRT; 34 AA.
ID Q70222
AC Q70222;

```



Matches 10; Conservative 7; Mismatches 16; Indels 11; Gaps 1

QY 346 TLPDNEVELKAFVAPAPPVETTTNYEWNLIHSPTDYQGEIKQGH 389  
 Db 2 TRPNNNTRKGVYIGP-----RAVFYATDIIGDIRQAH 34

RESULT 14

Q9Y181 PRELIMINARY; PRT; 27 AA.

ID Q9Y181;  
 AC Q9Y181;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Engrailed homeodomain protein (Fragment).  
 OS Priapulius caudatus.  
 OC Eukaryota; Metazoa; Priapulida; Priapulidae; Priapulidae; Priapulidae.  
 OX NCBI\_TaxID=37621;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA de Rosa R., Grenier J.G., Andreeva T., Cook C.E., Adoutte A., Akam M.,  
 RA Carroll S.B., Balavoine G.;  
 RA "Hox genes in Brachiopods and Priapulids: implications for Protostome  
 RT evolution.";  
 RL Nature 0:0-0(1999).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC EMBL; AF144895; AAD40651.1; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR009057; Homeodomain\_like.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR PROSITE; P850071; HOMEBOX 2; 1.  
 DR DNA-binding; Homeobox; Nuclear protein.  
 FT NON\_TER 1 1  
 FT NON\_TER 27 27  
 SQ SEQUENCE 27 AA; 3288 MW; 321AE16A0529F351 CRC64;

Query Match 0.8%; Score 45; DB 2; Length 27;  
 Best Local Similarity 50.0%; Pred. No. 4.7e+04;  
 Matches 13; Conservative 4; Mismatches 5; Indels 4; Gaps 2;

QY 829 LTEQRKDTLVRQLAVLLNVLDSIKV 854  
 Db 6 LTEQRQELAKDLA--LN--ESQIKI 27

RESULT 15

Q7R7E1 PRELIMINARY; PRT; 29 AA.

ID Q7R7E1;  
 AC Q7R7E1;  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Name=PY07646;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNL;  
 RC PubMed=12368865; DOI=10.1038/nature01099;  
 RX Carlson J.M., Angluoli S.V., Suh B.B., Kooij T.W., Perteza M.,  
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Seilengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,  
 RA Florens L., Yates P.R., Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,

RA Carucci D.J.;  
RT "Genome sequence and comparative analysis of the model rodent malaria  
RT parasite Plasmodium yoelii yoelii";  
EL Nature 419:512-519(2002).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AABL01002846; EAA20148.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 29 AA; 3752 MW; 3497968C69392CCB CRC64;

Query Match 0.8%; Score 45; DB 2; Length 29;  
Best Local Similarity 36.0%; Pred. No. 5.2e+04;  
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 935 LMWENLIORYIWDGSGNCEWSIFVY 959  
Db 1 LLLLEYFIKLYIWEYNNNNNNNAFYL 25

Search completed: October 12, 2005, 10:30:37  
Job time : 69 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	51	0.9	32	2	S55937	translation initia
2	50.5	0.9	25	2	I40692	cena protein (IGAI
3	49	0.9	25	2	B49111	POU homeodomain pr
4	47.5	0.9	27	1	TZAZ3	toxin III - snake-
5	44	0.8	34	2	S6635	tubulin alpha chai
6	44	0.8	35	1	N5GRPM	neurotoxin P2 - sc
7	42	0.8	22	2	S58433	isocitrate dehydro
8	41	0.7	25	1	SMWR	metallothionein -
9	41	0.7	29	1	Q1BP57	gene 1.5 protein -
10	41	0.7	35	2	F95172	hypothetical prote
11	40.5	0.7	28	2	C83797	hypothetical prote
12	40.5	0.7	31	2	D84404	hypothetical prote
13	40.5	0.7	31	2	T29628	hypothetical prote
14	40.5	0.7	31	2	G97596	hypothetical prote
15	40.5	0.7	34	2	I67427	retinoic acid rece
16	40	0.7	20	2	S63483	coenzyme P420 hydr
17	40	0.7	28	2	D49829	T-cell receptor va
18	40	0.7	29	2	G3690	neural cell adhesi
19	40	0.7	30	2	S12965	gamma-crystallin -
20	40	0.7	32	2	C49480	major immunophilin
21	40	0.7	33	2	I53186	ISI/cat protein -
22	40	0.7	35	2	JK0200	Gurmarin - Gymnema
23	39.5	0.7	23	2	A48186	ATP synthase beta-
24	39.5	0.7	28	2	S08569	ribosomal protein
25	39.5	0.7	30	2	A49312	cysteine proteinas
26	39.5	0.7	34	2	I67428	retinoic acid rece
27	39	0.7	18	2	PT0286	lg heavy chain CDR
28	39	0.7	20	2	A60372	pollen allergen Po
29	39	0.7	27	2	P00583	tyrosine 3-monooxy

Db 2 SVSTPPT---PSPSTP-----PTPSP 19

## RESULT 3

B49111

POU homeodomain protein (alternatively spliced) Oct-1B - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
C:Accession: B49111  
R:Das, G.; Herr, W.  
J. Biol. Chem. 268, 25026-25032, 1993  
A:Title: Enhanced activation of the human histone H2B promoter by an Oct-1 variant gene  
A:Reference number: A49111; MUID:94043371; PMID:8227066  
A:Contents: NTERA 2D1  
A:Accession: B49111  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-25 <DAS>  
A:Note: sequence extracted from NCBI backbone (NCBIP:139679)

Query Match 0.9%; Score 49; DB 2; Length 25;  
Best Local Similarity 34.5%; Pred. No. 3.6e+03;  
Matches 10; Conservative 5; Mismatches 10; Indels 4; Gaps 1;

QY 290 SPVLTVTPGSTSHSPTPTPTSAAPSESTP 318

Db 1 SPILAIFPSLSALQTTP-----PTTQP 25

## RESULT 4

TZAZ3

toxin III - snake-locks sea anemone  
C:Species: Anemonia sulcata (snake-locks sea anemone)  
C:Date: 30-Apr-1979 #sequence\_revision 24-Sep-1981 #text\_change 09-Jul-2004  
C:Accession: A91446; A91674; A01798  
R:Martinez, G.; Kopeyan, C.; Schweitz, H.; Lazdunski, M.  
FBBS Lett. 84, 247-252, 1977  
A:Title: Toxin III from Anemonia sulcata: primary structure.  
A:Reference number: A91446; MUID:78084776; PMID:23311  
A:Accession: A91446  
A:Molecule type: protein  
A:Residues: 1-27 <MAR>  
A:CROSS-references: UNIPROT:P01535  
R:Beress, L.; Wunderer, G.; Wächter, E.  
Hoppe-Seyler's Z. Physiol. Chem. 358, 985-988, 1977  
A:Title: Amino acid sequence of toxin III from Anemonia sulcata.  
A:Reference number: A91674; MUID:78044787; PMID:21843  
A:Accession: A91674  
A:Molecule type: protein  
A:Residues: 1-21, 'SC', 24-27 <BER>  
C:Comment: Three disulfide bonds are present.  
C:Superfamily: toxin III  
C:Keywords: venom

Query Match 0.9%; Score 47.5; DB 1; Length 27;  
Best Local Similarity 28.1%; Pred. No. 4.9e+03;  
Matches 9; Conservative 3; Mismatches 7; Indels 13; Gaps 1;

QY 60 ACCDLSSCDLAWFEGRCYLVSCPHKNCPEPK 91

Db 2 SCCP-----CYWGGCPWQNCYPE 20

## RESULT 5

C56635

tubulin alpha chain, brain-specific isotype (clone PTB6) - chum salmon (fragment)  
C:Species: Oncorhynchus keta (chum salmon)  
C:Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 10-Jul-1998  
C:Accession: C56635  
R:Coe, I.R.; Munro, R.; Sherwood, N.M.  
DNA Seq. 3, 257-262, 1992  
A:Title: Isolation of different brain-specific isotypes of alpha-tubulins from chum salmon  
A:Reference number: A56635; MUID:93208376; PMID:1296820

A:Contents: brain  
A:Accession: C56635  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-34 <COE>  
A:Note: sequence extracted from NCBI backbone (NCBIN:128392, NCBIP:128391)  
C:Superfamily: tubulin  
F:28/Binding site: polyglutamate (Glu) (covalent) #status predicted  
F:33-34/Cleavage site: Glu-Tyr (tubulin-specific carboxypeptidase) #status predicted  
F:33-34/Cross-link: peptide (Glu-Tyr) (by tubulin-tyrosine ligase) #status predicted

Query Match 0.8%; Score 44; DB 2; Length 34;  
Best Local Similarity 32.0%; Pred. No. 1e+04;  
Matches 8; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 151 EEMSEYXDDYRELEKDLLOPSGKQE 175

Db 6 EDMAALEKDYEEVGVDSEGESEEE 30

## RESULT 6

NFSRPM

neurotoxin P2 - scorpion (Androctonus mauretanicus)  
C:Species: Androctonus mauretanicus mauretanicus  
C:Date: 15-Oct-1982 #sequence\_revision 15-Oct-1982 #text\_change 09-Jul-2004  
C:Accession: A01758  
R:Rosso, J.P.; Rochat, H.  
Toxicol. 23, 113-125, 1985  
A:Title: Characterization of ten proteins from the venom of the Moroccan scorpion Androctonus mauretanicus  
A:Reference number: A94318; MUID:85193276; PMID:3992595  
A:Accession: A01758  
A:Molecule type: protein  
A:Residues: 1-35 <ROS>  
A:CROSS-references: UNIPROT:P01498  
C:Superfamily: scorpion neurotoxin  
C:Keywords: neurotoxin; venom  
F:1-18, 4-25, 15-30, 19-32/Disulfide bonds: #status predicted

Query Match 0.8%; Score 44; DB 1; Length 35;  
Best Local Similarity 30.4%; Pred. No. 1.1e+04;  
Matches 7; Conservative 5; Mismatches 9; Indels 2; Gaps 1;

QY 913 CLLKCSGHGHCPLTKRCICSHL 935

Db 15 CATCCGRGKC--VGPQCLCNRI 35

## RESULT 7

SS8433

isocitrate dehydrogenase (NAD) (EC 1.1.1.41) chain 2 - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 29-Nov-1995 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: SS8433  
R:Zeng, Y.; Weiss, C.; Yao, T.T.; Huang, J.; Siconolfi-Baez, L.; Hsu, P.; Rushbrook, J.I.  
Biochem. J. 310, 507-516, 1995  
A:Title: Isocitrate dehydrogenase from bovine heart: primary structure of subunit 3/4.  
A:Reference number: SS8432; MUID:95382766; PMID:7654189  
A:Accession: SS8433  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-22 <ZEN>  
A:CROSS-references: UNIPROT:Q7M3F1  
C:Superfamily: 3-isopropylmalate dehydrogenase  
C:Keywords: oxidoreductase

Query Match 0.8%; Score 42; DB 2; Length 22;  
Best Local Similarity 50.0%; Pred. No. 7.2e+03;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 528 PPVANAGPNHTITL 541

Db 7 PPSAKYGARHTVTM 20

```
RESULT 8
SMOR
metallothionein - cultivated mushroom
C:Species: Agaricus bisporus (cultivated mushroom)
C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 09-Jul-2004
C:Accession: A03286
R:Munger, K.; Lerch, K.
Biochemistry 24, 6751-6756, 1985
A:Title: Copper metallothionein from the fungus Agaricus bisporus: chemical and spectroscopic properties
A:Reference number: A03286
A:Accession: A03286
A:Molecule type: Protein
A:Residues: 1-25 <MUN>
A:Cross-references: UNIPROT:P04358
A:Experimental source: Mycelium; strain A-32
C:Comment: In contrast to the vertebrate metallothioneins, the fungal proteins bind copper
C:Comment: The absorptive, luminescent, and stereoptical properties of the copper MT are
C:Superfamily: metallothionein
C:Keywords: chelation; metal binding; metal-thiolate cluster

Query Match      0.7%; Score 41; DB 1; Length 25;
Best Local Similarity 41.2%; Pred. No. 9.7e+03;
Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 917 CSGHGHCDPLTKRCICS 933
Db 5 CSGASSCTCASGQCTCS 21

RESULT 9
QIBF57
Gene 1.5 protein - phase T7
C:Species: phase T7
C>Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 09-Jul-2004
C:Accession: I43002; I43004; S42294; A04417
R:Dunn, J.J.; Thompson, K.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94615
A:Accession: I43002
A:Molecule type: DNA
A:Residues: 1-29 <DUN>
A:Cross-references: UNIPROT:P03792
R:Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 148, 303-330, 1981
A:Title: Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of the right end
A:Reference number: A92866; MUID:82078034; PMID:7310871
A:Accession: I43004
A:Molecule type: DNA
A:Residues: 1-29 <DU2>
A:Cross-references: GB:V01127; NID:g15498; PIDN:CAA24338.1; PID:g15510
R:Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 166, 477-535, 1983
A:Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genes
A:Reference number: S42283; MUID:83241725; PMID:6864790
A:Accession: S42294
A:Molecule type: DNA
A:Residues: 1-29 <DUN>
A:Cross-references: EMBL:V01146; NID:g431187; PIDN:CAA24395.1; PID:g15574
C:Genetics:
A:Gene: 1.5
A:Map position: 19.48-19.69
C:Superfamily: phase T7 gene 1.5 protein

Query Match      0.7%; Score 41; DB 1; Length 29;
Best Local Similarity 46.7%; Pred. No. 1.2e+04;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 13 LLVITAGCAKQCSE 27
Db 7 LLVIVGCLAHCSD 21
```

```
RESULT 10
F95172
hypothetical protein SPI481 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: F95172
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidison, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, J.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95172
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-35 <KUR>
A:Cross-references: UNIPROT:Q97PV9; GB:AB005672; PIDN:AAK75575.1; PID:g14972972; GSPDB:G14972972
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI481

Query Match      0.7%; Score 41; DB 2; Length 35;
Best Local Similarity 42.1%; Pred. No. 1.6e+04;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 293 LVTVPGSTSTHSIPPTPSA 311
Db 17 VTILEGSHVLPSPYSA 35

RESULT 11
C83797
hypothetical protein BH1179 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: C83797
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirano, T.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and its reference number: A83650; MUID:20512582; PMID:11058132
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C83797
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-28 <STO>
A:Cross-references: UNIPROT:Q9KDN2; GB:AB001511; GB:BA000004; NID:g10173727; PIDN:BA0048173727
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1179

Query Match      0.7%; Score 40.5; DB 2; Length 28;
Best Local Similarity 38.1%; Pred. No. 1.2e+04;
Matches 8; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

Qy 957 FYVTVLAPTLIVLTGGTWC 977
Db 9 FALIVLVFLIVITGA-AWVC 28

RESULT 12
D84404
hypothetical protein Vng2544h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: D84404
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithaus, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon, J.; Hickey, E.K.; Holt, I.E.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: D84404
```

QY 958 YVTVLAFLLI---VLTGGFTWL 976  
|:||||: |: |  
db 8 YLTVLAISLLMWTIVILGVVYL 29

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:00:24 ; Search time 57 Seconds  
(without alignments)  
1403.926 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 5580  
Sequence: 1 MAPTGVLSLLLVITAGC.....VSMNGSIRNGASFSYCSKDR 1072

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	443	7.9	183	4	US-09-270-767-32456
2	312	5.6	1300	4	US-09-902-540-9932
3	215.5	3.9	827	4	US-09-248-796A-17307
4	206.5	3.7	372	4	US-09-902-540-10204
5	188	3.4	1093	4	US-09-902-540-13786
6	184.5	3.3	3969	3	US-08-061-376-5
7	184.5	3.3	3969	4	US-09-538-092-1262
8	183.5	3.3	822	2	US-08-474-067-9
9	183.5	3.3	822	2	US-08-474-068A-9
10	183.5	3.3	822	2	US-08-472-481-8
11	183.5	3.3	2870	4	US-09-478-467A-15
12	183.5	3.3	3178	4	US-09-479-467A-4
13	183	3.3	2254	4	US-09-949-016-9270
14	182.5	3.3	1020	4	US-09-434-123A-9
15	180	3.2	2736	4	US-09-252-991A-30227
16	179	3.2	880	4	US-09-902-540-10174
17	176.5	3.2	2231	1	US-08-153-799-16
18	176.5	3.2	3913	4	US-09-949-016-10933
19	176.5	3.2	4377	4	US-09-949-016-6978
20	173.5	3.1	1706	2	US-08-458-568-2
21	173.5	3.1	1706	2	US-08-399-411-2
22	173.5	3.1	1706	3	US-08-516-859A-2
23	173.5	3.1	1706	3	US-09-586-472-2
24	173.5	3.1	1706	4	US-09-528-706-2
25	173	3.1	2315	4	US-09-543-681A-5434
26	171.5	3.1	1810	4	US-08-793-273C-4
27	171.5	3.1	1810	5	PCT-US95-11684-4

28 170.5 3.1 1861 2 US-08-790-912-4 Sequence 4, Appli  
29 168.5 3.0 4302 3 US-08-658-136-5 Sequence 5, Appli  
30 168.5 3.0 4303 2 US-08-460-751-2 Sequence 2, Appli  
31 168.5 3.0 4303 4 US-09-479-467A-2 Sequence 2, Appli  
32 168 3.0 4302 3 US-09-052-469-8 Sequence 8, Appli  
33 168 3.0 4302 4 US-08-422-582-8 Sequence 8, Appli  
34 168 3.0 4302 4 US-09-052-262-8 Sequence 8, Appli  
35 168 3.0 4339 3 US-09-052-469-6 Sequence 6, Appli  
36 168 3.0 4339 4 US-08-422-582-6 Sequence 6, Appli  
37 168 3.0 4339 4 US-09-052-262-6 Sequence 10208, A  
38 167.5 3.0 385 4 US-09-902-540-11984 Sequence 11984, A  
39 166.5 3.0 991 4 US-09-902-540-11984 Sequence 5503, Ap  
40 166.5 3.0 3892 4 US-09-328-352-5503 Sequence 1378, Ap  
41 164.5 2.9 610 4 US-09-538-092-1378 Sequence 4, Appli  
42 163 2.9 1596 4 US-08-978-277A-4 Sequence 2, Appli  
43 162.5 2.9 1439 2 US-08-449-644-2 Sequence 2, Appli  
44 162.5 2.9 1439 2 US-08-087-244A-2 Sequence 4, Appli  
45 162 2.9 1266 3 US-08-506-296B-4 Sequence 4, Appli

## ALIGNMENTS

## RESULT 1

US-09-270-767-32456

; Sequence 32456, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 32456

; LENGTH: 183

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-32456

Query Match 7.9%; Score 443; DB 4; Length 183;

Best Local Similarity 50.3%; Pred. No. 1.8e-27;

Matches 92; Conservative 22; Mismatches 69; Indels 0; Gaps 0;

Qy 473 INGPFIIEKTSVDSFVLRLSNLDPGNYSFRITVDSGATNSTTAALIVNNAVDPYPVAN 532

Db 1 ISGPIGYQVLPVNTLQLDLTSGNTFKLITVDSNNVTSTTATIAVLKETYAPVAN 60

Qy 533 AGPNHTITLPONSLTNGNOSDDHQVLVWLSLGPSEGHVWVGQVTPYLHLSAQOE 592

Db 61 AGDAVILYLPNNVTLNGTASSDDHEIVAWEWTKDASDEAKAVDMQNTTRTPYVQLSNLEE 120

Qy 593 GDYTFQLKVTDSRQOSTAXVTIVQENRPPVAVAGPKELIFFVESATLDGSSSSDD 652

Db 121 GMYTFVLKVTDSGQOSTAKVHVFKPPTNSPPVAREAGSTTTSLPINWLLNGSDSKDD 180

Qy 653 HGI 655

Db 181 IGI 183

## RESULT 2

US-09-902-540-9932

; Sequence 9932, Application US/09902540

; Patent No. 6833447

; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Wiegand, Roger C.

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

; FILE REFERENCE: 38-10(15949)B



; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 9932  
; LENGTH: 1300  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-9932

Query Match 5.6%; Score 312; DB 4; Length 1300;  
Best Local Similarity 23.9%; Pred. No. 1.5e-15;  
Matches 184; Conservative 98; Mismatches 291; Indels 198; Gaps 41;

QY 155 EYXDDYR-----ELEDKLLQPSKQEPGRGSAEYTDWGLLPGSEGAFFNSVGDSPAPVPAETQ 210  
DB 588 DFEDDYDGAIVREISD-----GGQTVWDIG-----DPIHNSVLEIYTPG----- 625  
QY 211 QDPHELHYNESATPA-----PKLPERSVLLPLPTTPSSGEVLEKEKASQLQEQSSNSG 265  
DB 626 -----NLNPLAKPALSSYNQFQGM-----LPATIDLTAYAGKTVQVRVGTDRSG 674  
QY 266 KEV--LMPSHSLPPASLESLSVTVKSPVLTVTPGSTSHSIPPTPSAPSES--TPSEL-- 321  
DB 675 GYTGMLLD-----DLEFNIGT--NTPPATIAPEDGVCNWPPIANAGPDRSIAPGELV 725  
QY 322 -----PISPT-----TAPRTVKE-----LTVSAG-- 340  
DB 726 SLYGSAADPEGQFVFTTWAQTSQGPVTLAGANTLNPSTAPQVTESTALVFTLTVSDGVK 785  
QY 341 ----DNLIIT--LPDNEVELKAFVAPAPPVETTVN-----YEMNLISHPTDYQGB- 384  
DB 786 TSTDSTVTVALPNN-----PP-----TVNAGLDGIVERAEYTLGSSASDADGNA 831  
QY 385 -----IKQGHKQTLNLSQSLVGL---YVFKVTYSSENAFGEFVNVTVKPA 427  
DB 832 LTVLMTQVSGTPVAVKDYTPPTATFPAPEVTLDESIVFRLTVSDGIVATVNDVTVTVNA 891  
QY 428 RRNLPPVAVSPQQLTLPLTSALIDGQSTDDTEIVSYHWEENGPFIEKTSVDSP 487  
DB 892 NRA--PIVSDTSVAFAGTVTITASAVD-----PGDALSYSWEQTGGSTV-AINGADTS 943  
QY 488 VLRLSNLDPCNTSFRLTVDSDGATNSTTAALIVNNAV--DYPPVANAGPNHTITLPPONS 545  
DB 944 ALSFATVPGSEYFTVATDGSASAKAVPVTIIDGSLPANAPTVMAGIDATAN-AGDV 1002  
QY 546 ITLNGNQSSDDHQIVLYEW--SLPGSEGKHVVMQGVQTPYLHLSAMQEGDYT---FQLKV 601  
DB 1003 VTILGSSASDAEGDTLYHWEQIG---GTDVTLTGADTLTPSTAPSTASGTTLGFILTV 1058  
QY 602 TDSRROQSTAXTVIVVOPE-----NNRPPVAVAGPKELIFPVESATLDCSSSDDHGIV 656  
DB 1059 SDGTSTSTSDA--VRVIVAADPGANPSNTAPEADAG--ESAIVAEGATVTLNGTATDADGDTL 1116  
QY 657 FYHWEHVRGPSAVEMENIDKAIATVVG-LQVGYTHPRLTAKQOGLSSTSL---TVAV 711  
DB 1117 VIWVTOI--GGTPVTLSDASLTPTFTAPASSDPLTFLMVS-----GRTVVDVTSISV 1170  
QY 712 KKENNSPPRARAGRHVLVLPNNSTIIDGSRSTD--DQIRIVSYLWIRDGQSPAAGVDIGS 770  
DB 1171 TEENVAPV---ASARAVLSGNQTSATLDGSASSANGDVLTYRWT---QVSGFNATISGA 1224  
QY 771 DHSVALQLNLVE-----GVYTHLRLVTDQGSASDTHDTATVEVQPPRKG 816  
DB 1225 DQAVA--VNLPLDLDKTAASFSPRLTVKOSAGTESSTT-----VQFSARNGG 1269

## RESULT 3

US-09-248-796A-17307  
; Sequence 17307, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17307  
; LENGTH: 827  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17307

Query Match 3.9%; Score 215.5; DB 4; Length 827;  
Best Local Similarity 23.3%; Pred. No. 3.8e-08;  
Matches 168; Conservative 77; Mismatches 314; Indels 161; Gaps 35;

QY 171 SGKQEPGRGSAEYTDWGLLPGSEGAFFNSVGDSPAPVPAETQODP---ELHYNESAS---- 223  
DB 172 TSKETPTSTNAQT-----SSANNQOQSNTAAPSTSVIOPSTSEVHVQSQOFTSTPN 223  
QY 224 --TPAPKLPERSVLLP-----LPTTPSSGEVLEKEKASQLQEQSSNSGKEVLMP 271  
DB 224 TPTSPNTPTTSEAPPTTSAAPTTSAPPTTSEAPVTSTSEVVENTPTT---SEAPNT-----P 273  
QY 272 SHSLPPASLESLSVTVKSPVLTVTPG--STEHSIPTPTSAAPSESTPSELPISTPA- 328  
DB 274 TTSEAPVTPSTSEV--VPNTPTTSKAPNTPTTSEAPATPTTSEAPNTPTTSEAPVPTTISE 332  
QY 329 --PRT-----VKELTVSAGDNLI---ITLPDNEVELKAFVAPAP---PVETT 367  
DB 333 VVPTTSTQGDVSTSTSVTEQTTLTSTLPTTALTQTSTPEASDPKPSSTSIETP 392  
QY 368 YNYEWNLI SHPTDYQGEIKQGHKQTLNLSQSLVGLYVFKVTYSSENAFGEFVNVTVKPA 427  
DB 393 STSTFEQDPTTSSVGTSPSSSEQOPQPTTSELA-----VTSNSPTQESTSLVEFTSSL 445  
QY 428 RRNLPPVAVSPQQLTLPLTSALIDGQSTDDTEIVSYHWEENGPFIEKTSVDSP 487  
DB 446 ESSNTP---TENPSTSE-AQPSTGA---SQAPPD-----TSSAPAP 480  
QY 488 VLRLSNLDPCNYSFR-----LVTDSGDGATNSTTAALIVNNAVDYPPVANAGPNHTIT 540  
DB 481 ELSSNADFSNLHLHSSSETTSLVNPTDSDQIDSSSTTDA--VSQATTEPTSENT-PTAASS 537  
QY 541 LPQNSITL-----NGNQSSDDHQIVLYEWSLPGSEGKHVVMQGVQTPYLHLSAMQEGD 594  
DB 538 VTANDINSAQYSAPTSNADAEATASSPVSEQLATGSGTSLDTTACASSTASEATA--ENL 595  
QY 595 YTFQLKVTDSRROQSTAXTVIVQENNRPP-----PVAVAGPKELIFPV--ESAT--L 644  
DB 596 STFG---TDGSSDAS---QTIAETTSNSPQSVVTPSASASPDVSTL-PTGSESGTSLV 647  
QY 645 DGSSSSDDHGIVFYHWEHVRGPSAV-EMENI-----DKAIAITVTLQVCTVHF 691  
DB 648 SGSETSIDNTV-----ASGSTVIPESNIPTQSPQSVSVSSDAANVSTGSAITDGL 701  
QY 692 RLTVKDXQOGLSSTSLT--VAVKKENNSPPRARAGRHVLVLPNN---SITLDGSRSTD 746  
DB 702 AGSETGVQPISSATGTSEPFSSSEYNS-----SEGTSLVVTPTNSELSSVTGSEATA 756  
QY 747 QRIVSYLWIRDGQSPAAGVDIDGSDHSHVALQLTNLVEGVYTHLRLVTDQGSASDTHDTATV 806  
DB 757 TAINSESVLTGSSDTAA--TVTGSILTGNTSETSATAIASESTLTGTGTGATDSAAATTI 814

## RESULT 4

US-09-902-540-10204  
; Sequence 10204, Application US/09902540

	: Patent No. 6833447	
	: GENERAL INFORMATION:	
	: APPLICANT: Goldman, Barry S.	
	: APPLICANT: Hinkle, Gregory J.	
	: APPLICANT: Slater, Steven C.	
	: APPLICANT: Wiegand, Roger C.	
	: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof	
	: FILE REFERENCE: 38-10(15849)B	
	: CURRENT APPLICATION NUMBER: US/09/902,540	
	: PRIOR FILING DATE: 2001-07-10	
	: PRIOR APPLICATION NUMBER: 60/217,883	
	: PRIOR FILING DATE: 2000-07-10	
	: NUMBER OF SEQ ID NOS: 16825	
	: SEQ ID NO 10204	
	: LENGTH: 372	
	: TYPE: PRF	
	: ORGANISM: Myxococcus xanthus	
	: US-09-902-540-10204	
	Query Match                3.7%; Score 206.5; DB 4; Length 372;	
	Best Local Similarity     25.1%; Pred. No. 5.3e+08;	
	Matches      86; Conservative    58; Mismatches 144; Indels    55; Gaps    16;	
Qy	410 SSNAFGGPNVTVKPARVNLPVPVVAVSPQ-----LQLTLPLTSA 452	
Dd	:	:
Dd	18 TDAAGGGFWIDIAFTGLNTPTALVDQQCTTGTHVDAQGEDLADELVS---V 73	
Qy	453 LIDGSOSTDTEIVSYHWEINGPPI-----EETSDPSVLRLSLNDPGNSFRLTVT 506	
Dd	:	:
Dd	74 QLQSGSSLAGGALTVMQEAGPPAAMSGADATATPSTAPEVLADM--LTERLVLI 129	
Qy	507 DSGGNSTTAALIYNNAVDPYPAVANAGPHNITILPONSI TLNGH-QSDDHQI VLYEWS 565	
Dd	:	:
Dd	130 DG-GLRDRDTVQLVRQ-VNKAPAVASGAQTVD-EGSVTTLOQSAEPDGDAVVSQQMT 186	
Qy	566 LGPSEGEKVMMGVQ--TPYLHLSAMQG--DYTFOLKVTDSRSRQSQSTAXVTVIVOPEN 621	
Dd	:	:
Dd	187 ---QVSGLPVTLTGADTLTPSTAPAVAKGSBELVFLVLSOGQLSGEAATVSVTR--- 240	
Qy	622 NRPEVA--VAGPDKElifPVESATLDGSSSDHDHGIFVFYHMEHVGRGPSAVEMENTDKAIA 679	
Dd	:	:
Dd	241 -HYPLAPTYSAGDDVATFSASKSVTLTATDEQGTILTAYMSQSSEGPT-VTIQNATAQRV 298	
Qy	680 TVTGLQGT---XHFRLTKVKOQGLSSTSTLTVA VKKENNSPP 719	
Dd	:	:
Dd	299 QFTAPEVSAATEFVVKKVSSGESLGSAEDRVTVT---SPAAPP 338	
	RESULT 5	
US-	-09-902-540-13786	
	: Sequence 13786, Application US/09902540	
	: Patent No. 6833447	
	: GENERAL INFORMATION:	
	: APPLICANT: Goldman, Barry S.	
	: APPLICANT: Hinkle, Gregory J.	
	: APPLICANT: Slater, Steven C.	
	: APPLICANT: Wiegand, Roger C.	
	: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof	
	: FILE REFERENCE: 38-10(15849)B	
	: CURRENT APPLICATION NUMBER: US/09/902,540	
	: PRIOR FILING DATE: 2001-07-10	
	: PRIOR APPLICATION NUMBER: 60/217,883	
	: PRIOR FILING DATE: 2000-07-10	
	: NUMBER OF SEQ ID NOS: 16825	
	: SEQ ID NO 13786	
	: LENGTH: 1093	
	: TYPE: PRF	
	: ORGANISM: Myxococcus xanthus	
	: US-09-902-540-13786	
	Query Match                3.4%; Score 188; DB 4; Length 1093;	
	Best Local Similarity     21.0%; Pred. No. 9.8e+06;	
	Matches      180; Conservative    102; Mismatches 323; Indels    254; Gaps    42;	



Db 3057 SQISNAVOTPPHLPKATEKLVVNNQMQBLYVLQTPNGVTKIQITSSVSSTPSVME 3116  
Qy 367 TY-----NYENLISHPTDYQGEIKQGHKQTL-----N 394  
Db 3117 TNTSVLPGMGGLTTLTCLNPSLPTQSGLFSPASGLLPMSSHQHLHSPFAATQSSPPPN 3176  
Qy 395 LSQLSVGLYV-----FKVTVSSENAFGEFVNVT-----KPARRV----- 430  
Db 3177 ISNPPSGLLIGVQPPDPQLLVSESSQRT-----DLSTTVATPSSGLKRPISRLQTRKNK 3232  
Qy 431 -----NLPPVAVVSPQ-LQELT---LPLTSALID-GSQSDTDDTEIVSYHWEENGP 476  
Db 3233 KLAPSSPTSINAPSDVSNWTLINFTPSQLPNHPSLLDGLSLNTSSHRTV-----P 3283  
Qy 477 FI--BEKTSV-----DSPVL-----RLSNLDPGNYSFRLTVDSDGA 511  
Db 3284 NIKRKSSIMVFEPAPLLPQSVGTAATAAGTSTISQDTSHLTSGSVS-----GL 3334  
Qy 512 TNSTTAALIVNNAVDPYPPVANAG-PNH-TITLPO-----NSITLNGNQS-----DD 556  
Db 3335 ASSSVLNVVSMQTTTPTSSASVFGHVTLTNPRLLGTPDIGSISNLLIKASQOSLGIQD 3394  
Qy 557 HQIVLYEWSLPGSGEGKHVVMQGVTPYVHLSAMQEGDYTFQLKVYDSSRQGSTAXVTVI 616  
Db 3395 QPVAL-----PPSGMPPQLGTSQTP-----STAAITAAASIC 3427  
Qy 617 VQPENRPPVAVAGDKELIPFVESATLDGSSDDHGIVFYHWEHVRG--PSAVEMENI 674  
Db 3428 VLPSTQTTGITAASP-----SGEADH-----YQLQHVNQLLASKTGHS 3468  
Qy 675 DKAIATVTVGLQVTHFRLTVKDOGLSSTLTVAVKENNSPPRARAGRHVLVLPNN 734  
Db 3469 QRDLDSAGPOVSNFTQTVADAPNSMGLRQNKALSSAVQASPTSP-----CG-----SP 3516  
Qy 735 SITLDGSRSTDORIVSVLWIRDGOSPAAG-----DVIDGSDHSHVALQL 778  
Db 3517 SSPSGORS-----SPSPVGTFTKPKTKRFPQLDKNGKHKHVS--- 3558  
Qy 779 TNLVEGVYTHFRLVTDGQ-----ASDITDITAVEVQDPDKSG 816  
Db 3559 -----HLRTSSSEAHIPDQETTSLSGTGTPGAEEQDITASVE-QSSQKEG 3605

## RESULT 8

US-08-474-067-9  
; Sequence 9, Application US/08474067  
; Patent No. 5811518  
; GENERAL INFORMATION:  
; APPLICANT: Ranscht, Barbara  
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/474,067  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/213,361  
; FILING DATE: 14-MAY-1994  
; APPLICATION NUMBER: US 07/607,293

; FILING DATE: 30-OCT-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1682  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 822 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-474-067-9  
  
Query Match 3.3%; Score 183.5; DB 2; Length 822;  
Best Local Similarity 21.4%; Pred. No. 1.4e-05;  
Matches 165; Conservative 90; Mismatches 260; Indels 257; Gaps 40;  
  
Qy 359 APAPPVETTYNYE-----WNLSHPTDYQGEIK-----QGHKQTLNLSQLSVGLYFKVTVSS 411  
Db 142 ADSPP-EGVFTIEKESGMLLHMLPDLREKIVKYELYGHAVSENGASVEEPMNI-SIIIVTD 199  
Qy 412 ENAFGEFVNVTVPARRVNLPPVAVVSPOLQELTLPPLTSALIDGSDTDD-----TEIV 466  
Db 200 QDNKPKFTQDTFRGS-----VIEGMFGTSMQVMTATDEDDAVNTYNGVV 245  
Qy 467 SY--HWBEINGPF-----IBEKTSVDSFVLRLSNLDPGNY-SFRLTV--TDSGATNSTT 516  
Db 246 AYSIHSQEPKEPHDLMTFIHKSTGTISVI--SSGLDREKVPYRLTVQATDMDCGSGTTT 303  
Qy 517 AALIV-----NNAVTPPVANAGPNHTITLPQNSITLNGNQSDDHQIVLYEWSLPGGS 570  
Db 304 AEAVVQI LDANDNAPEFEP-----QKYEAWVPENEV-----GH 336  
Qy 571 EGKHVVMQGVQTP-----YLHLSAMQEGDYTFQLKVYDSSRQGSTAXVTVIQVPPENR 623  
Db 337 EVQRLTVDLQVNPWPAWRATYTHIVGGDDGQH-----FTTTHPETNQ 379  
Qy 624 PPVAVAGPDKELIPFVESATLDGSSDDHGIVFYHWEHVRGPGSAVEMENIDKAIATVTG 683  
Db 380 --GVLTKKGLDF-----EAQDQH--TLYVEVTNEAPFAVKLP----- 413  
Qy 684 LQVGYTHFRLTVKDOGLSSTLTVAVKENNSPPRARAGRHVLVLPNNISITLDGSR 743  
Db 414 -----TATATVVHVVKOVNEAP-----VFVPPSKVIEAQEGIS 446  
Qy 744 -----TDDORIVSVLWIRD-----GQSPAAGVIDGSDHSHVALQL 778  
Db 447 IGEVLCTYTAQDPDKEDQKI-SYITISRDPAWMLAVDPSGQITAAAG-ILDRDSQF----- 500  
Qy 779 TNLVEGVYTHFRLVTDGQASDITDITAVEVQDPDKSGELVELTLQVGVGQLTEQKQDITLV 838  
Db 501 --VKNNVYEVWVLATDSGNPPTTGTGL-----LLTLTDINDHGPIPEPRQIIIC 548  
Qy 839 RQLAV--LLNVLDSDIKVQKIRAHSDLSVIVFYVQSRPPPKVLKAAFAVARNLHMRLSKE 896  
Db 549 NQSPVPQVLNITDKDLSFNS-----SPFOALQTHSDSDIYMAEVS-----E 589  
Qy 897 KADFL-----LFKVLAVDTAGCLLKCSGHCHCDPLT--KRCICS-HLWMENLQRYWDGES 950  
Db 590 KGDVTVALSKKFLKQDVTYDLHLSLDHGNREQLTWIRATVCDCHGQVNDQCPR-PWKGG- 647  
Qy 951 NCEWSIFVTVLAFTLVLTGGFTWLCICCCCKRQKTKIRKTKYITLIDNDEQERMLR 1010  
Db 648 -----FILPILGAVLALLT---LLALLLVLRKR-----KVKEPLLLPEDD----- 686  
Qy 1011 PKYGIKHRSTHNSLMVSESEFSDQDTIFSRKMERG-----NPKVSMNGSI 1059  
Db 687 -----TRDNVFFYYGEGGGBEDQD--YDITQLHRLGLEARPEVVLRLNDV 727

RESULT 9





```
; PRIOR APPLICATION NUMBER: 60/115,127
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3178
; TYPE: PRT
; ORGANISM: C. Elegans Lov-1 protein
US-09-479-467A-4

Query Match      3.3%; Score 183.5; DB 4; Length 3178;
Best Local Similarity 19.4%; Pred. No. 0.00013;
Matches 147; Conservative 97; Mismatches 259; Indels 253; Gaps 29;

QY 195 FNSVGDSPAVPAETQQQDELHYLNESASTAPKLPERSVLLPLPTTPSSGVL----- 248
Db 327 FTSTASTSTSTTQQ-----SSSTTSPSSTTSTSTPTTTTPTTSTLSSLPDN 379
QY 249 -----EKEKASQLOEQSSNSGKEVLMPSHSL-----PP 277
Db 380 AICSYLDETTSTTTTMTLSTTTEEPSTSTTTTEVTSTSTSTVTTTPTTTLTTSTAST 439
QY 278 ASLELSSVTVEKSP-----VLTVPGSTEHSIPTPTSAAPSESTPSELPIPTTAP--- 329
Db 440 STTEPSTVTWTPSPSTPVTSTVSSSSSTVTTTPTSTESTSTSPSTVTTSTTAPSTS 499
QY 330 -----RTVKELTVSAGNLIITLDPNEVE 353
Db 500 TTGPSSSSSTPSTASSVSSSTASSTQSTSTQSSSTTKSTSTSSDG---TNPDPFV 556
QY 354 LKAFVA--PAPPVETTYNEMNLI SHPT-----DYQGEIKQGHKQTLNLSQ----- 397
Db 557 EKATTFYDSTSVNLTLNSGLGIIGVQTSIECTSPSTSSNYVSTTKDGCFTKSVSMPLG 616
QY 398 -----LSVGLYVFKVTSS-----NAGEGFVNVTVPARRVNLPPVAVVS- 439
Db 617 GTPYASTFVGPNGYTFRATMTDDKKVYTYANVYQIYSSTTIESGSTS-----AVASS 672
QY 440 ----PQLQELTLPLTSALIDGS--OSTDDTEIVSVHWBEI--NGPFTTEKTSVDSPLRLS 492
Db 673 TSTSTSPSTLS--TSVTPESTSTSSDSTTTSAGSTTTLQESTTTSSESTDSSTTIS 731
QY 493 NLDPGNYSFRLVTSDGATNSTAALIVNNAVDYPPVA--NAGPNHTITLQNSITLN-- 549
Db 732 DTSTSTSPSTADSTSTLSVDQDFILDGLSMWNETRHEDSNIVPLPTNAITPTER 791
QY 550 -----GNQSSDQHIVL-----YWSLGPSEGHVVMQGVQTPYLHLSAMQSGDYTF 597
Db 792 SQTFCRNVTSTPFLIKESTCLNYSNTVLNATYSSNIPIETFLVGI-----GTVEF 846
QY 598 QLKVTDSRQQTAXVTVIVQPNENRPPVAVAGPKELIFFVESATLDGSSSSDDHGIVF 657
Db 847 RINWDTLTMQVSHIFTL-----NVVA--DSTSTSEVTSTSTGSSE----- 888
QY 658 YHWEHVRGSPAVEMENIDKAIATVTGLQVGYHFLRTLVKDDQGLSSTSTLTAV--KKEN 715
Db 889 -----SSAISTSGIE-----STSTLEASTTSDASD 914
QY 716 NSPPRACGRHVLVLPNNSITLDGSRSTDORIVSYLWIRDGQSPAGDVIDGSDHVA 775
Db 915 STSTSDSG-----TTSSTTIDSNST-----PSTSD-----SSG 945
QY 776 LQLTNLVEGYTFHLRVTDQSASDTDATVEVPD 811
Db 946 LSQT-----PSDSSSASDS-MRTTVDPD 968
```

## RESULT 13

```
US-09-949-016-9270
; Sequence 9270, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9270
; LENGTH: 2254
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9270

Query Match      3.3%; Score 183; DB 4; Length 2254;
Best Local Similarity 20.8%; Pred. No. 8.2e-05;
Matches 174; Conservative 83; Mismatches 302; Indels 276; Gaps 37;

QY 170 PSGQOEPRGS-----AEYTDNGLLPGS-----EGAFNS-----SVGDSPAVP 206
Db 1325 PERKEEPPRPFLBQPLLGELTVGTVPDLSLWSLVAQGFDFSMVQYKDAQGQPAVP 1384
QY 207 AETQQ-----DPELHY-----LNESAS-----TPA 226
Db 1385 VAGDENVTVPGLDPDRKIKNLYGLRQRQRPESVVAKTAPOEDVDETSPBTLGHEA 1444
QY 227 PKLPERSVLLPLTPSPSGEVLE-----KEKASQLOEQSSNSGKEVLM 271
Db 1445 PESPEPLLGELTVTGSSPDSLSLFTVPQGSFDSFTVQYKDRGRPRAVRVGKSEVT 1504
QY 272 SHSLPPA-----SLELSSVTVEKSPVLTVTQGSTHSIPTPTPTSAAPSESTSELPI 323
Db 1505 VGLGEPGHKYMHLGLHEGQRPVSAVGVTAPQOEE-----TPATESPLEPRIGELTV 1560
QY 324 SPTTAPRTVKELTVSAG--DNLIITLPNEVELKAFVAPAPVE--TTNYEWNLSHPD 380
Db 1561 TDVTNSVGLSWTVPGEQFDSFIQYKDKQGPQVVFVAADQREVTVNLE----- 1611
QY 381 YQGEIKQGHKQTLNLSQLSVGLYVFKVTVSSENAFGEFVNVTVPARRVNLPPVAVVSP 440
Db 1612 -----PERKYMNYGLHDG-----QRMGPLSVVITVAPATEASKPP---LEP 1651
QY 441 QLQELTLPPLTSALIDGOSTDDT--EIVSYHWEENINGPIEIKTSVDSPLRLSLNDPGNY 499
Db 1652 RLGELTV-----TDITPDSVGLSWTVPGEF-----D 1678
QY 500 SFLRTVTDSDGATNSTTAALIVNNAVDYPPVANAGPNHTITLQNSITLNGNQSSDDHQI 559
Db 1679 SFVQYKDRDQOQ-----VVPVAAD-----QREVTIPDLERKRYK 1716
QY 560 VLYEWSLG-----PGSEGHVVMQGVQTPYLHLSAMQEGDYTFQLKVTDSRQQTAXVTV 615
Db 1717 LLFGIQDGKRSPSVSVEAKTVARGDASPG---APPLG-----ELWVTDPDPSLRLSWTV 1769
QY 616 IVQPENNRPPVAVAGPKELIFFVESATLDGSSSSDDHGIVFYHWEHVRGSPAVEMENID 675
Db 1770 ---PEGQFDSFVQFKDXD-----GPQVFPVEGHE 1796
QY 676 KAIATVTGLQVG--TVHREL-----TVKDDQGLSSTSTLTITVAVKKENNSP 718
Db 1797 RSV--TVPDLDGRKYRFLLYGLLKKRKGPLTADGTTTEARGAMDDTGT-----KRP 1846
QY 719 PRARAGGR--HVLVLPNNSITLDGS--RSTDDQRIVSYLWIRDGQS---PAAGDV-----ID 768
Db 1847 PKPRLGEELQVTTVTQNSVGLSWTVPEQGFDSFVQYK--DRDQGPQVVPVEGSLREVSV 1905
QY 769 GSD--HSVALQITNLVEGYTFHL--RVTDQSASDTDATVEVPDPRKSLVELTLQVG 825
Db 769 GSD--HSVALQITNLVEGYTFHL--RVTDQSASDTDATVEVPDPRKSLVELTLQVG 825
```





Db 1463 GQVTADGSGN-----WSFTFG-----IPLPDGTVVNVVARSP 1494  
QY 604 SSRQOSTAXVTV-----IVQENRRPPVAVAGPKELIF-----PVESATLDGSSS 649  
Db 1495 SNVDSAPAVITVDGVAPAPVIDPSNGTEISGTAEAGATVILTDGGGNPIGOATADGSG- 1553  
QY 650 SDDHGIVFYHWEHVRGSAVEMENIDKAIATVTGLQVGYHFRFTVKDQOG-LSSTSTLT 708  
Db 1554 -----NWTFTFG-----TPLANGTV-INAVAQDPAGNTSGPASVT 1587  
QY 709 VAVKKENNSPPPARAGGRHVLVLPNNSITLDGSRSTDDQRIVSYLWIRDGOSPAAGDV-I 767  
Db 1588 V-----DAIAPAP-----VINPSNGVVISGTAEAGATVILT-----DGNGNPICQVTA 1631  
QY 768 DGS DSHVALQLTNLVEGVYTHLRTVTDGASD TDT-ATVE-VQP-----DPRKSGLVEL 820  
Db 1632 DSGGKWAFTPATPLANGT-VINALAQDAAGNNSPTSATVDSLAPAAPVIDPSNGSVIAG 1690  
QY 821 TLQVG-----VGQLT 830  
Db 1691 TAEAGATVILTDGNGNPICQVT 1712

Search completed: October 12, 2005, 10:19:17  
Job time : 62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2005, 10:00:24 ; Search time 52 Seconds  
(without alignments)  
1983.546 Million cell updates/sec

Title: US-10-764-390-3  
Perfect score: 5580  
Sequence: 1 MAPPTGVLSLLLVITAGC.....VSMNGSIENGASFYSCKSDR 1072  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225	4.0	26926	1 I38344	titin, cardiac mus
2	208.5	3.7	6805	2 S20901	titin - rabbit (fr
3	208	3.7	5291	2 F90696	hypothetical prote
4	205	3.7	1742	2 S24600	projectin - fruit
5	203.5	3.6	5188	2 B85547	probable RTX fami
6	202.5	3.6	6658	2 T13931	projectin - fruit
7	201.5	3.6	3283	2 AC1018	large repetitive p
8	200.5	3.6	849	2 D82510	chitinase VCA0027
9	200.5	3.6	1009	2 C64483	hypothetical prote
10	195.5	3.5	1020	2 A46405	C protein alpha an
11	195.5	3.5	2232	2 T34434	hypothetical prote
12	192.5	3.4	1000	2 I46521	titin - rabbit (fr
13	190.5	3.4	1461	2 E90696	hypothetical prote
14	190.5	3.4	1461	2 A85547	hypothetical prote
15	190	3.4	3083	2 AH2493	hypothetical prote
16	188	3.4	528	2 I47141	gastric mucin (clo
17	188	3.4	4006	2 T09070	probable tenascin
18	187.5	3.4	3507	2 T34513	hypothetical prote
19	187	3.4	2013	2 A11489	probable peptidogl
20	187	3.4	3624	2 AD0835	large repetitive p
21	186	3.3	2944	2 A54849	collagen alpha 1(V
22	184.5	3.3	822	1 IUMSCP	P-cadherin precurs
23	183	3.3	6831	2 A88852	protein unc-22 (im
24	183	3.3	6839	2 S57242	twitchin [similar
25	183	3.3	7160	2 T27935	hypothetical prote
26	182.5	3.3	1275	2 T33369	hypothetical prote
27	181.5	3.3	3968	2 A44265	trithorax homolog
28	179	3.2	5175	2 T20992	hypothetical prote
29	179	3.2	5198	2 T43290	hemikenttin precurs

RESULT 1

I38344  
titin, cardiac muscle [validated] - human  
N/Alternate names: Connectin  
N/Contains: serine/threonine-specific protein kinase (EC 2.7.1.1--)  
C/Species: Homo sapiens (man)  
C/Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004  
C/Accession: I38344; I38345; S20897; S20898; S20899; S63665; S37393  
R/Labeit, S.; Kolmerer, B.  
Science 270, 293-296, 1995  
A/Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
A/Reference number: A57430; MUID:96026330; PMID:7569978  
A/Accession: I38344  
A/Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL  
A/Molecule type: mRNA  
A/Residues: 1-26926 <LAB>  
A/Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:G1017424; PID:G1017425  
R/Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.  
Biochemistry 34, 553-561, 1995  
A/Title: Dissecting titin into its structural motifs: identification of an alpha-helix m  
A/Reference number: I38345; MUID:95119041; PMID:7819249  
A/Accession: I38345  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1977-2014 <MUS>  
A/Cross-references: EMBL:X83270; NID:G602579; PIDN:CAA58243.1; PID:G602580  
R/Labeit, S.; Gaucel, M.; Lakey, A.; Trinick, J.  
EMBO J. 11, 1711-1716, 1992  
A/Title: Towards a molecular understanding of titin.  
A/Reference number: S20897; MUID:92258380; PMID:1582406  
A/Accession: S20898  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 13597-14200, 'I', 14202-14696 <LAB>  
A/Cross-references: EMBL:X64698; NID:G37192; PIDN:CAA45939.1; PID:G37193  
A/Accession: S20897  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB>  
A/Cross-references: EMBL:X64699; NID:G37190; PIDN:CAA45940.1; PID:G37191  
A/Accession: S20899  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2  
A/Cross-references: EMBL:X64697; NID:G37190; PIDN:CAA45938.1; PID:G37195  
R/Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.  
J. Mol. Biol. 256, 556-563, 1996  
A/Title: Genomic organization of M line titin and its tissue-specific expression in two  
A/Reference number: S63665; MUID:96177761; PMID:8604138  
A/Accession: S63665  
A/Status: nucleic acid sequence not shown

heparan sulfate pr  
probable translati  
glucan 1,4-alpha-g  
zonadhesin - mouse  
tenascin-X - bovin  
ankyrin 3, long sp  
hypothetical prote  
probable oligopept  
peptidoglycan link  
zinc finger protei  
probable peptidogl  
tenascin-X precurs  
proline-rich pepti  
projectin - fruit  
hypothetical prote  
fibronectin - bovi

[illegible]

A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin.  
A:Reference number: I46520; MUID:90238553; PMID:2129545  
A:Accession: I46520  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 4235-5250 <LA2>  
A:Cross-references: EMBL:X17329; NID:q1756; PIDN:CAA35207.1; PID:g930251  
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro  
C:Keywords: muscle

```
Query Match      3.7%; Score 208.5; DB 2; Length 6805;
Best Local Similarity 18.8%; Pred. No. 0.029;
Matches 212; Conservative 155; Mismatches 363; Indels 395; Gaps 59;

Qy 52 PPVVDCTAACCDLS-----SCDLA---WFFEGRCYLVSCHPKENCEP 90
Dy 861 FDISDIDADACSLSHIPLDGGSNITNYIVEKCDVSRGDW-----VTALASVTKTSCRI 915

Qy 91 KMGPIRSYLTFLVR-----PVORPAQLLDYGDWMLNRGSPSGIWDGSPEDIR--- 138
Dy 916 GKLIPOGEYV-FRVAENRFGISEPLQSPKMAQF-----PFGV-PSEPKNARVTK 964

Qy 139 --KDLXFLGKDW-----GLEEMSEYKDDYRELEKDLLOPSGKQEPGSAEYTDMLPGSE 192
Dy 965 VNKDCIFVANDRPDSDGSGSPITGYLIE-RKGRNSLLWVKANDTAVRSTYPCAGLVEGLE 1023

Qy 193 GAFNSSVGDSPAPVPAETQQDPPELHYLNESASTPAPKLPERSVLLPLPT-TPSSGEVLE-- 249
Dy 1024 YSF-----RIVALNKAAGSSP-PSKPTVEYVARTVPDPKPEVIDVMT 1064

Qy 250 KEKASQLQEQSSNSGKEVL-----MPS-----HSLP-----PASLELSVVT 286
Dy 1065 KSTVSLIWARPKHGGSKIIGFYVEACKLPQDKWVRNCTTTHQIPHEEYVTGLEENAOY 1124

Qy 287 VEKSPVLTVTPGSTEHSITPTPTSAAPSESPSELPTPTAPRTVKELTVSAGDNLIIIT 346
Dy 1125 QFRAIAKTAVNISQPSSELRTVTHA--ENVPPRIDIS--VAMKSL--LTVKAGTNVCL- 1177

Qy 347 LPDNEVELKAFVAPVPPVETTYNYEWNLISHPTDYQCEI---KQGHQTLNLSQSLVGLY 403
Dy 1178 -----DATVFGKMPVTV-----SWK-----KEGTVLKPAEGIKWAMQORNLCTLELF 1218

Qy 404 VFK-----VTVSSENAFGEFVNVTVPKARRVNLPPVAVV-----SPQLQE 444
Dy 1219 SVNRKSGDYITITAESSGSKSATIKLVLDRLPG-PPASVKINRWYSDRAMLSPEPLED 1277

Qy 445 LTLELTALIDGSDTDDTEIVSVHWEIEINGPFIEEKTSVDSPVLRLS---NLDPGNYSF 501
Dy 1278 GGSEITNIVDKRETS-----SRNAAQV-----SANVITSCSVEKLIHGHEYQF 1322

Qy 502 RLTVTDSGATNST-TAALI VNNAVDPYPVANAGPNHHTITLPNSITLNGQSSDD---- 556
Dy 1323 RICAENKYGVDPVFTBPAIAKN--PYDPPGRCDDPPVLSNVTKDHMTVSWKPPADGGSP 1380

Qy 557 -----HQVLVWLSLGPSEGKHVVMQGVQTPYLHLSAMQEG-DYTFOLKVYDTS 604
Dy 1381 ITGYLLEKRETHAV---NWT---KVNRPVITERTIKA-----TGLQEGTEYEFRTAINK 1429

Qy 605 S--RQOSTAXTVIVQPNENRPPVAVAGDPKELIFPV-----ESATLDGSSSDHGH-- 654
Dy 1430 AGPKPSDASKANVAQ-----DPLYPGPP---APPKVVDTTRSSVLSWKGKPAYDGGSP 1481

Qy 655 IVFY-----HWEHVRGFSAVEMENIDKAIATVTGLQVGT-YHFRLTVKQOQGLSS 703
Dy 1482 IIGYLVEKTRADTNWRCNLP-----QKLQKTFEVTGLMENTEYQFVAVNVKGVSD 1536

Qy 704 TS-----TLTVAVKKNNGSP----- 719
Dy 1537 PSDVPDKHCPKDIILIPPEGLDADLRKTLILRAGVTNRLYVVPVGRPPPKITMSKPNVNL 1596

Qy 720 PARAG-----GRHVILVLPNN----- 734
Dy 1597 RERIGLDIKSTDFDFTFLRCENVNKYDAGKITLTLENSCGKGYTIVVKVLDTPGPPVNV 1656
```

```
Qy 735 --SITLDGSRSTDDQRIVSVLWIRDGQSPAAGDVIDGSD-----HSVALQLT 779
Dy 1657 VKEISRDSAYITWDPPIV-----DQSGSPIINYVVEKRDADARKSWSTVTTTECSKTSFVS 1710

Qy 780 NLVEG-VYTFHLKRVDSQG-----ASDQTDATATVEVO-----PDPR 813
Dy 1711 NLEEGSKYFRVVAENEYGIQDGETDAVKASSETPGVVDLKVLTVTKSSCNIGWKKEP 1770

Qy 814 KSGVELTLOVGVGQLTEQRK-DTLVRQLAVLNLVDS-----IKVKIRAHSDLS 864
Dy 1771 SDGSRITGYV-VDFLTEENKQWVKMSLSLOYSTKDLNEGKQYTFRVSANENGEGTPTS 1829

Qy 865 TVIVFYVQSPPPKVYLKAAEVAARNLHNR-----LSKEKADPLL 902
Dy 1830 EITV-----VAKDDVVPADLDLKDLPDLCLYAKENSFRL 1864
```

## RESULT 3

F90696  
hypothetical protein EC0542 [imported] - Escherichia coli (strain O157:H7, substrain R1)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: F90696  
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F90696  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5291 <HAY>  
A:Cross-references: UNIPROT:Q8X2T1; GB:BA000007; PIDN:BAB33965.1; PID:g13360000; GSPDB:G  
A:Experimental source: strain O157:H7, substrain R1MD 050952  
C:Genetics:  
A:Gene: EC0542

Query Match 3.7%; Score 208; DB 2; Length 5291;  
Best Local Similarity 20.6%; Pred. No. 0.021;  
Matches 173; Conservative 114; Mismatches 319; Indels 234; Gaps 34;

```
Qy 190 GSEGAFFNSSVGDSPAPVPAETQQDPPELHYLNESASTPAPKLPERSVLLPLPTPSSGEVLE 249
Dy 2854 GSDLAL-SGTSDDQPAITAIT---VTLNGQYNSATTDAS--GNWSEVTPASAVSALGEATY 2907

Qy 250 KEKASQLQEQSSNSGKEVLMPGSHLPPASLESSTVVEKSPVLTVTPGSTEHSIPTPT 309
Dy 2908 SVTASVTNAQNSGSTASHNVQVNTALPGITIN----PVATDDIINASEAGSAQTISGQVT 2963

Qy 310 SAAPSESTPSELFIPTTAPRTVKELTVSAGDNLIIITLP-----DNEVELKAFVA-- 359
Dy 2964 GAAAGSVTVVLELGGKTYTA-----TVQADLSNNVSVPAADWQALNGELTVNASVTNA 3016

Qy 360 -----PAPPVETTYNVE-WNLISHPTDYQCEIKQG----- 388
Dy 3017 VNTGSGTRDITIDASLPLGLRVDTVAGDDVVNIIIEHA---QAQVITSGSGFAAGTALT 3073

Qy 389 --HKQTLNLSQSLVGLVFKVTVSSENAFGEGFVNVTVPKARRVNLPPVAVVSPQLQELT 446
Dy 3074 VINNQTVAAVTLANGSVSGVGPATDVSNWPAGTLNITVSGANSAG-TQTSITHP----LT 3128

Qy 447 LPLTSALIDGQSQDSD-----TETVSYHWE-----EING 475
Dy 3129 VDLTAVAI SMNSITSDDAINAAEKGAAALTLGSGTSGVEAGQTVTVTGGKTYTTTVAANG 3188

Qy 476 PFTEKTSVSDSPVLRLNLPNGYNSFRLTWTDSGATNSITAA-----LIVNAV 525
Dy 3189 SW-----STTVPAADLAALRDGDASAQVRVTVNNG--NSATATHEYSVDSAAPTVTINTI 3241

Qy 526 DYPPVANA-----GPNHTITLPQNSITLNGQSSDDHQLVLY---EWSLGPQSGEG 572
Dy 3242 ASDNIINASEAAAGVTVSGTSTAQTGQTLIVTLNGT-----NYQTTVTQDGSWSL----- 3291
```

```
QY 573 KHVMMQGVQTPYLHLSAMQEGDYTFQLKVTDSRQOSTAXVTVIQVPENNRPPVA---VA 629
Db 3292 -----TLPASDITALANNGYTLTATVSDLAGNLSASKGVTV--DTTAPVIFSNIVA 3341
QY 630 GPD-----KELFFVESATLDGSSSDDHGIVFYHWEHVRGVSAMENIDKAIATVGL 684
Db 3342 GDDVNNVHEHIQAQIISGTATGAVAGDRLVVTIAGQVVTSTDASGNMWSGVGPASVISGL 3401
QY 685 QVGTVHFRLTWKDQGLSTST-----LTVAVKKNNSPPRARAGRHVLVLPN 733
Db 3402 ADGVTVISATITDSAGNSSTQHNQVNTAAVLSVSTISGDNLLNAAEAG----- 3452
QY 734 NSITLDGSRSTDQRIVSYLWIRDGQSPAAGDVI----DGS DHSVALQLTNLVEGVYTFH 789
Db 3453 SALTLSGT-----GTNFATGIVTVLLNGKYSATIO-----SNGSWSVN 3492
QY 790 LRVTPDQASGADTDTATVEVQDPDRKSLVELTLQVGQOLTQRKDTLVRQLAVLLNV-- 847
Db 3493 VPAADVAALS DGTSVTSASAQD-----SAGNGNSSTQTHNVQVNTAAVLSVST 3542
QY 848 LDSDIKVQKIRAHSDLS-----TVIVFY-----VQSRPPFKV-LKAAEVA 886
Db 3543 TSGDNLLNAAEAGSALTLSGTCTNFATGTVTVTVLLNGKYSATIOQNSGWSVNVPAADVA 3602

RESULT 4
S24600
projectin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C:Accession: S24600
R:Fyherg, E.
submitted to the EMBL Data Library, May 1992
A:Reference number: S24600
A:Accession: S24600
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1742 <FVR>
A:Cross-references: UNIPROT:Q24463; EMBL:X66018; NID:g8377; PID:g8378
C:Genetics:
A:Gene: FlyBase:bt
A:Cross-references: FlyBase:FBgn0005666

Query Match 3.7%; Score 205; DB 2; Length 1742;
Best Local Similarity 19.6%; Pred. No. 0.0062;
Matches 187; Conservative 99; Mismatches 258; Indels 410; Gaps 48;

QY 80 VSCPHKENC-----EPKMG--PIRSYLTFLVLRVQRPQAQLLDYGDMLNRGSPSGIWD 132
Db 133 VSEVHKDGGCKLWKPKDGGEPVESYLVKEFDP-----DTGIW-- 171
QY 133 SPEDIRKDLXFLGKDWGLEEMSEYXDYRELEKDLLQPSGKQEPGRGSABYTDWGLLPGE 192
Db 172 -----LPVGRSD---GPEYNDGLVPGHD 192
QY 193 GAFN-----SSVGDSPAVPAET-----QQDPDLHYLINESASTPA-PKLPEPS-----V 234
Db 193 YKFRVAVKAVNKEGESE--PLETLGSIIAKDP-----FSVPTKGPVPEPTDWTNRKVEL 242
QY 235 LLPLPT-----TPSSGEVLE-KEKASQLQPSQSSSGKEVLMPSHSLPPASLESLSVTVEK 289
Db 243 ANPEPASDGGSPFIQGIYVEVKYKYSPLWEK-----ALETNSPT--- 280
QY 290 SPLVTV-----TPGSTHSIPTPTPSAAPSESTFSELPISPTTAPRTV 332
Db 281 -PTATVOGLIEGNEYQFRVVALNKGGLSEPSDKIFTAKPY-----LAPKIDRRNL 332
QY 333 KELTVSAGDNLIITLPDNEVELKAFVAPAPPVETTYNNYEMNLI SHPTDYQGRIKQGHKT 392
Db 333 RNTIUSSTGAL-----KLDANITGEPAPKV-----EWKUSNY-----HLQSGKNVT 373
QY 393 LNLSQLSVGLYV-----FKVTVSSNAFGEFVNVTV--KPARRVNLPPVAVVSP 440
```

```
Db 374 IETPDYTKLVIRPTQSRSDSGEYLVTATNTSGKDSVLNVVVITDKPS-----PP-----NG 424
QY 441 QLQELTLPLTSALIDGQSSTD--TEIVSYHWEI---NGPPIEEKTSVDSPLRLSLND 495
Db 425 PLQISDVHKGCHLWKAPSDGGGTPLEYFQIDKLEPETGCIWPSCRSTE-PQVDVTGLS 483
QY 496 PGN-YSFRLTVTSDGATNSTTA--ALIVNNAVDPYPVANAGPNHTITLPQNSITLNGNQ 552
Db 484 PGNEYKFRVSAVNAEGESQPLVGDESIVARNPFDEP-----GKPN---LKATD 529
QY 553 SSDDH-----QIVLY-----EWSLG---PGSEGHVVMQGV--QTPY 584
Db 530 WDKDHVLDLAWTPPLIDGGSPICYIIIEKQDKYKWERALDVPADQCKATIPDLVEGQTYK 589
QY 585 LHLSAMQEG-----:||||----- 593
Db 590 FRVSAVNAAGTGPSPDSTPIIAKARNKPIIDRSSILVEVRIKAGQSFPPCKVSGEPAP 649
QY 594 -----DYTFOLKVTTDSSRQOS-----TAXVTV 615
Db 650 QTKWLLKKEVYKDNVKVTNVDTYNTKLKVN SATRSDSGIYTVFRENANGEDSADVKTV 709
QY 616 IVQENNRPPVAVAGPKELIFFVESATLDGSSSDDHGIVFYHWEHVRGVSAMENID 675
Db 710 IDKP---APP---NGPLKVDEINSECTLHWNPPDDGG-----QPID 746
QY 676 KAI-----ATVTGLQVG-----TYHFLTVKDOOGLSSTSLTVAV 711
Db 747 NYVVKGLDETTGRWMTAGETDGPVTALKVGLTLPGHKYKFRVRAKNAGSEPLTTAAQI 806
QY 712 KKENNSPPRARAGRHVLVLPNNSITLDGSRSTD--DQRIVSYLWIR---DQQSPAAGDV 766
Db 807 IAKN-----PFDVPTKPGTPIKDFKDFVDLEWTRPEADGGSPITGV 850
QY 767 IDGSDH-----SVALQLTN-----LVEGV-YTFHLRVTDGSGASDITDTAT 805
Db 851 VEKRDKFSPOWEKCAEISDDITNAHVPLDIEGLKYEFVRVAVNKAAGPGSPDAT 904

RESULT 5
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: B85547
R:Perina, N.F.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Groetbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: UNIPROT:Q8X4H5; GB:AE005174; NID:g12513369; PIDN:AAG54838.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615

Query Match 3.6%; Score 203.5; DB 2; Length 5188;
Best Local Similarity 20.8%; Pred. No. 0.036;
Matches 165; Conservative 109; Mismatches 296; Indels 223; Gaps 33;

QY 190 GSEGAFFNSVGDSPAVPAETQDDPELHYLINESASTPAPKLPERSVLLPLPTTPSSGEVLE 249
Db 2854 GSDLAL-SGTSQDPAGTAIT---VTLNGQNTSATTTDAS--GNWSVTVPASAVSALGEATY 2907
QY 250 KEKASQLQEQSSNSGKEVLMPSHSLPPASLESLSVTVEKSPVLTVPSTGSTEHSIPTPT 309
Db 2908 SVTASVTNQAQNSSTASINNVQNTALPGIITN-----PVATDDIINASAGSAQTISGQVT 2963
QY 310 SAAPSESTPSELPISPTTAPRTVKELTVSAGDNLIITLP-----DNEVELKAFVA-- 359
```





R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Karpman, A.; White, N.; Farrar, M.; et al. (2003) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Paratyphi A. *Nature* 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; et al.  
A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Paratyphi A. *Nature* 413, 848-852, 2001  
A;Reference number: AB0502; PMID:21534947; PMID:11677608  
A;Accession: AC1018  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-3283 <PAR>  
A;Cross-references: PAR:AL513382; PIDN:CAD09244.1; PID:gl6505248; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY4458

Query Match	3.6%;	Score 201.5;	DB 2;	Length 3283;
Best Local Similarity	20.5%;	Pred. No. 0.024;		
Matches 163;	Conservative 99;	Mismatches 245;	Indels 289;	Gaps 38;
Qy	271	PSHSLPAPLSSLVTE	-----KSPVLVTTPGSTEHSITPPPTSAAPSEST-----PSE	320
Db	2547	PTFLGNIDADARYVTVEVGHGKVELTKATGATWSVTPTGTWADGGYTLTVRVEDD	2606	
Qy	321	LPISPTTAPRTV--KELTVSA	-----GDNLI-----ITLPD--NEVEL---	354
Db	2607	AGNVKYSAPLTVTVDTQITDIVIELVNDNGIPGDNLTNDYRPHPRVTPGDNVNEVRLSID	2666	
Qy	355	--KAFVAPAPVETTYNEM	-----NLISHPTDYQGEIKOGHKQTTNLSQLSVGLY	403
Db	2667	GGNTWVRATOGCTAGIMDYTPWPKVDVTLGLHTLTVEATD-----KAGNKTQTOLD	2714	
Qy	404	VFKVTVSSENAFGBGVNVTVPKARRVNLPPVAVVS	PQLQELTLPLTSLALIDGSQSDDTT	463
Db	2715	---FTIUT-----RUSTPTIAM-----DSRDDT	2734	
Qy	464	EIVSYHWEINGPPIEBKTSVDSVPLRLSNLD	-----495	
Db	2735	GAIGDH-----ITSVKRPGFTIGNIDADAHSVILLRITOGGNSQEVLTTOVGQW	2783	
Qy	496	-----PGNYSFRLTVTDS	-----DGATNSTTAALIVNNAVDPYPVAN	532
Db	2784	RFTPDADWAGSGYTLTVEVDINDAGNRQSTPLVVTVDVTQTSITDITLVNDHGVPDDNLTN	2843	
Qy	533	A-GPNHTITLPO--NSITLGNQSSDDHQIVLYEM-SLGPSECKHVVMQGVOTPYLHLS	588	
Db	2844	STREQFEITPADVNSVQLSDIGGAN-----WVSATQGIEG-----VMGYTWP-----T	2897	
Qy	589	AMQEGDYTFQKLVDSRSRQOSTAXTVITVQPPNNRPVAVAGRPDKELIFPVESATLDGSS	648	
Db	2888	DMGDKHITLVMTVDRAAGTATQTLFFIDRLSTPIAL-----GLQV-----	686	
Qy	649	SSDDHGIVFYHWEHVRGVSAAVEMENIKAIATVT-----	686	
Db	2929	STDDTGTGPDGMTNRTPTFTFI-LQNIQSDVLNVTVSVTHNGTTTSFTATQAGAGWSFTPP	2987	
Qy	687	-----GYVHFLTVKQOQGLSSTST-LTVAVKKE-----NNSPPRABAGGRHV---	728	
Db	2988	APWGDGYTLTVTVEDRAGNTRPSTPLTVTVDTQTQIAIDHIELVNDSGVPGDNTVKHVRPQ	3047	
Qy	729	--LVLPNN-----SITLDGSRSTDDQRIVSYLVLIWRIDGOSPAAAGVDIDGSDHVALQTLNLV	782	
Db	3048	FQISVPDDVEKVLISIDGGTI-----WVTAIKSSTVG--IWDYT-----WPTDMP	3050	
Qy	783	EGVYTFHLRVTDSGASDPTDA--TVEVQPPDRKSGLVELTLQVGVGQLTEQRKD-----TL	837	
Db	3091	EGOHTLIVEVDGAGNKMKTGLDFTIDT-----LLTPTIELAPDQDTQGNKNDNLTS	3143	
Qy	838	VRQLAVLLNVLDSPIKQKIRAHSDLSIVIVFYQSRPFPFKVLKAAAEVARNLHRLSKEK	897	
Db	3144	VTQPVFVLGSIKQVR-----HVELS-----IEHNGTFKTVVLTSADGWRVRPDSAL	3191	
Qy	898	AD-ELLFKVLRVDTAG	912	

Df 3192 ADCSYTFVTIVTDVAG 3207

RESULT 8  
D82510  
chitinase VCA0027 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C;Accession: D82510  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson,  
J.; Ermlaeva, M.D.; Vanathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Song,  
L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae  
A;Reference number: A82035; UID:20406833; PMID:10952301  
A;Accession: D82510  
A>Status: preliminary  
Molecule type: DNA  
A;Residues: 1-849 <HEI>  
A;Cross-references: UNIPROT.O9KND8; GB:AEO04346; GB:AEO03853; NID:g9657401; PIR:  
A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
C;Genetics:  
A;Gene: VCA0027  
A;Map position: 2  
C;Superfamily: chitinase with C-terminal chitin-binding domain

Query Match 3.6%; Score 20.5; DB 2; Length 849;  
Best Local Similarity 23.8%; Pred No. 0.0038;  
Matches 112; Conservative 67; Mismatches 187; Indels 105; Gaps 23

Qy 367 TYN--EW-NLISHPT-----DYQEIKQHKKQTILNLS---QLSVGL 402  
Db 390 TYDFYGNWNVLGHOTALYCSPFMRQCCKGVDENGEPKYGPAYTTDNIGLIQLLAQG 449

Qy 403 YPKVTYSSENAEGGFVNITVKPARVNLPPVAVVSFLQELTLP--LTSLD---- 455  
Db 450 PPCKLVVGAA-MYGRGWGEVTPASLKDPDMFTGVNGKLGTTAQGVWEAGVIDYKG 508

Qy 456 ---GSOST-----DDTEIVSYHWEELNGPFI--EEKTSVDSPLRLSNLD-PGNY 499  
Db 509 NFMLGANKTGVNGFEYCDQAEPWNRTTGVLTFDDDRSKAKGAVRNLAGLF 568

Qy 500 SFRLTVTSD-----GATNSTTAALI VNNAVDPVPVANAGPHHTITLQNSITLNQ 552  
Db 569 SWEIDADNGDLNAMHGELAGGTTTP-----PVNKAPVANAGADITVTGPA-AVSLDGSA 622

Qy 553 SSD-DHOI VLVEYSLGPSGKGVVMQGVOTPFYLHLASA---MQEEDTYFLQKVDTSSRQQ 608  
Db 623 SKSDSGIASYLEW---QTAGPATVL TGANSKAKASFNAAEVETKQFTFFKLTVTDNKAT 679

Qy 609 STAXVVVIPOE-----NNRPVAVAGDKELIFPVESATLDGSSSD-DHGI V FYHWH 662  
Db 680 ATDPVVTVVNP KSTTPNTAPAVALSAPAS--VKAGATAVVDASGDADDQDPLSFMTWDL 737

Qy 663 VRGSAVEMENIDKAIAITVTCLOV-----GTYHFRLTVKDQQCLGSTITLTVAVK 712  
Db 738 PVGVN-----ATVOGAQVFVAGEVTQDTTLDTFVTVSDGKATS KASAVLVEK 786

Qy 713 KENNPSPRARAGRRHVLPN-NSITILDGSRSTDQRIVSYLVIRIDGSPA 762

bh 787 K-----ATTGGDA CTNIWNAFSI VTGCOCVWTAGKTWEAKWTRTGPDPS 830

RESULT 9  
C64483  
hypothetical protein MJ1468 - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: C64483  
F:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996





Db 810 ITKDSVTLHWDLPIDGGSRAITNYIVKREATRKSYSVTYTKCHKCYKVTGLSEGCYF 869  
Qy 691 FRLTVKDDQOGLSSTSTLTVAVKKENSPPRARAGRHVLVLPNNISITLDGSRSTDQRIV 750  
Db 870 FRVMAENEYIGIBSETPKEPKVASEAPSP-----PDSLNIMDTIKST-----V 912  
Qy 751 SYLWIR---DQSPAAQDVID-----GSD-----HVALQ-----LTNLVEG-VYTFHLRVTD 794  
Db 913 SLAMPKPHDGGSKITGYVIEAQKSGDQWTHITTVKGLCVVRNLTEGEYTFQVMAVN 972  
Qy 795 SQGAS 799  
Db 973 SAGRS 977

RESULT 13  
E90696  
hypothetical protein ECs0541 [imported] - Escherichia coli (strain O157:H7, substrain R1  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C;Accession: E90696  
R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: E90696  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1461 <HAY>  
A;Cross-references: UNIPROT:Q8XD19; GB:BA000007; PIDN:BA033964.1; PID:g13359999; GSPDB:G  
A;Experimental source: strain O157:H7, substrain R1MD 050952  
C;Genetics:  
A;Gene: ECs0541

Query Match 3.4%; Score 190.5; DB 2; Length 1461;  
Best Local Similarity 21.1%; Pred. No. 0.029;  
Matches 139; Conservative 81; Mismatches 214; Indels 225; Gaps 28;  
Qy 191 SEG--AFNSSVGDSPAPVPAETQDPPELHLYNESASTPAPKLPERSVLLPLPTTPSSGEVL 248  
Db 442 AEGYAVNASVSDRAG-----NTTSHSANFTVDTSPVVSNTV-----AGDDIL 486  
Qy 249 EKEKASQLQEQSSNSGKE-----VLMPSHSLPPASLESLSSVTVEKSPVLTVTPGSTE 302  
Db 487 NNAEQVAQIIISGVSGASPGDVTYVKGTHVLGTGIVLADGWNVALDPAVTRILDRGAN 546  
Qy 303 SIPTPTSAAPSETPSE-----LPISPTAPRTVKELTVSAGDNLITLDPNEVELKAFV 358  
Db 547 TIFVTVTTAAAGNTGAASRAITLVGVSPLIITNTV-----SGDDII-----SGA 589  
Qy 359 APAPPVETTYNWEWLISHPTDYQGEIKQGHKQTLNLSQSLVGLYVFKVTVSSENAFGE 418  
Db 590 EKGAPLTLTGSTQ-----QAEQTGQTVTVTLAGOS-----FTTTVQADGWSW--- 629  
Qy 419 FVNVTVKPARVNLPPVAVVSQQLQELTLPLTSALIDGSOSTDTEIVSVHWEINGPFI 478  
Db 630 --SLTVPAAMGNLPGDGA-----ITASVTLDSGNTGNT----- 662  
Qy 479 EEKTSVDSPLRLNLDPGNYSFRLTVTDSGATNSTTAALIVNNA---VDYP--PVANA 533  
Db 663 SRTITVDSQAPALS-IDP-----LTADNIIAASGGQDLPTGTGTTDA 703  
Qy 534 GPNHTITLQNSITLNGQSSDDHQIVLYEWSLPGSGEGKHVMQGVQTPVLHLSAMQEG 593  
Db 704 QPGQGTVTTLNGQTYQGVVQPDG-----TWS-----VTVPAANVGALADG 743  
Qy 594 DYTFLQKVTDSRRQSQSTAXVTIVIQENNRPPVAVAGP---DKELIFP-----VESATL 644  
Db 744 NATVTASVNDVAGNPSSVSRLV---DATPPVVTINPVDNTVINTPEHAQAQIIISGTV 800  
Qy 645 DGSSSSDDHGIVFYHWEHVRGSPSAVEMENIDKA-----IATVTGLQVGTYH 690  
Db 744 NATVTASVNDVAGNPSSVSRLV---DATPPVVTINPVDNTVINTPEHAQAQIIISGTV 800  
Qy 645 DGSSSSDDHGIVFYHWEHVRGSPSAVEMENIDKA-----IATVTGLQVGTYH 690

Db 801 TGAQAGD---IV-----TVTLLNNVDYTTVVDGSGNWSLGVPSVVSGLADGSYP 846  
Qy 691 FRLTVKDDQOGLSSTSTLTVAVKKENSPPRARAGRHVLVLPNNISITLDGSRSTDQRIV 750  
Db 847 VSVSVTDKAGNTGQSLSLTVV---NTAAP-----LIGINSI----- 879  
Qy 751 SYLWIRDDQSPAGDVIDGSDHVALQLTNLVEGVYTFHLRVTDSCASDSTDITATVEVQ 809  
Db 880 -----AGDDVINASEKGADLQITG-----TSDQPVNTAITVTLNGQ 915

RESULT 14  
A85547  
hypothetical protein Z0609 [imported] - Escherichia coli (strain O157:H7, substrain EDL9  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: A85547  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouais, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: A85547  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1461 <STO>  
A;Cross-references: UNIPROT:Q8XD19; GB:AB005174; NID:g12513364; PIDN:AAG54837.1; GSPDB:G  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z0609

Query Match 3.4%; Score 190.5; DB 2; Length 1461;  
Best Local Similarity 21.1%; Pred. No. 0.029;  
Matches 139; Conservative 81; Mismatches 214; Indels 225; Gaps 28;  
Qy 191 SEG--AFNSSVGDSPAPVPAETQDPPELHLYNESASTPAPKLPERSVLLPLPTTPSSGEVL 248  
Db 442 AEGYAVNASVSDRAG-----NTTSHSANFTVDTSPVVSNTV-----AGDDIL 486  
Qy 249 EKEKASQLQEQSSNSGKE-----VLMPSHSLPPASLESLSSVTVEKSPVLTVTPGSTE 302  
Db 487 NNAEQVAQIIISGVSGASPGDVTYVKGTHVLGTGIVLADGWNVALDPAVTRILDRGAN 546  
Qy 303 SIPTPTSAAPSETPSE-----LPISPTAPRTVKELTVSAGDNLITLDPNEVELKAFV 358  
Db 547 TIFVTVTTAAAGNTGAASRAITLVGVSPLIITNTV-----SGDDII-----SGA 589  
Qy 359 APAPPVETTYNWEWLISHPTDYQGEIKQGHKQTLNLSQSLVGLYVFKVTVSSENAFGE 418  
Db 590 EKGAPLTLTGSTQ-----QAEQTGQTVTVTLAGOS-----FTTTVQADGWSW--- 629  
Qy 419 FVNVTVKPARVNLPPVAVVSQQLQELTLPLTSALIDGSOSTDTEIVSVHWEINGPFI 478  
Db 630 --SLTVPAAMGNLPGDGA-----ITASVTLDSGNTGNT----- 662  
Qy 479 EEKTSVDSPLRLNLDPGNYSFRLTVTDSGATNSTTAALIVNNA---VDYP--PVANA 533  
Db 663 SRTITVDSQAPALS-IDP-----LTADNIIAASGGQDLPTGTGTTDA 703  
Qy 534 GPNHTITLQNSITLNGQSSDDHQIVLYEWSLPGSGEGKHVMQGVQTPVLHLSAMQEG 593  
Db 704 QPGQGTVTTLNGQTYQGVVQPDG-----TWS-----VTVPAANVGALADG 743  
Qy 594 DYTFLQKVTDSRRQSQSTAXVTIVIQENNRPPVAVAGP---DKELIFP-----VESATL 644  
Db 744 NATVTASVNDVAGNPSSVSRLV---DATPPVVTINPVDNTVINTPEHAQAQIIISGTV 800  
Qy 645 DGSSSSDDHGIVFYHWEHVRGSPSAVEMENIDKA-----IATVTGLQVGTYH 690  
Db 801 TGAQAGD---IV-----TVTLLNNVDYTTVVDGSGNWSLGVPSVVSGLADGSYP 846  
Qy 691 FRLTVKDDQOGLSSTSTLTVAVKKENSPPRARAGRHVLVLPNNISITLDGSRSTDQRIV 750

```

Db      847 VSVSVTDKAGNTGQSLSLTVTV---NTAAP-----LIGINSI----- 879
Qy      751 SYLWTRDQSPAGVVDGSDHSHVALQLTNLVEGVYTHRLRVDSQASDTHDTATVEVQ 809
Db      880 -----AGDDVINASEKAGDLQITG-----TSDQPVNTAITVTILINGQ 915

RESULT 15
AH2493
hypothetical protein all7128 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AH2493
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
  Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
  DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
  A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3083 <KUR>
A:Cross-references: UNIPROT:O8YL10; GB:BA000020; PIDN:BAW78212.1; PID:gi17135666; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7128
A:Genome: plasmid

Query Match          3.4%; Score 190; DB 2; Length 3083;
Best Local Similarity 21.1%; Pred. No. 0.091;
Matches 167; Conservative 96; Mismatches 302; Indels 226; Gaps 35;

Qy      148 WGLEMSXYD--DYRELEKDLQPSGKQPRGSAEYTDWG----- 186
Db      1976 YGSLALSSYVDIGVRYSNL--DL-----ATGSGEISTWGPVHRYSNIERFDITGTR 2025
Qy      187 -----LLPGSEGAFFNSVG---DSPAVPAET-----QDDPE-----LHYLE 220
Db      2026 YGDELLGGNLDKLTGGGADTLKAGLGNDDTVILAAQTAGGSKIEDGDGENDTLDLTINL 2085
Qy      221 SASTPA-----PKLPERSVLLPLP-----TTPSSGEVL-----EKEKASOLQEQSNSSG 265
Db      2086 SLSTFTIGTAGIQRLL-GYTLILDLDNQDGIITPESDLSIINFNSSAGTGFIKVDNLSG 2144
Qy      266 KEVLMPSHSLPASLESLSVTVKSPVLTPFCSTHSHIPTPTSAAPSESTPSELPISP 325
Db      2145 TDIL---NKLFGNSANQAPVT-QANKVLTV---AEDSVTTPLAIATPTDNDLLITITI 2196
Qy      326 TPAPRTVKELTVSAGDNLLIITLPDNEVELKAFVAPAPPVETTYNYEWNLI SHPTDYQGEI 385
Db      2197 TAVPEASKG-----IIRLPDNTV-----VTVN----- 2218
Qy      386 KQCHKQTLNLSLSGLYVFKVTVSSENAFGE-GFVNVTVPKARKVNLPPVAVVSPOLQE 444
Db      2219 -----TLLTQQLT-----SLVFVSVVNVANGSAGSFYTVSDGK-----GTASQT 2259
Qy      445 LTLPLTSA-----LIDGSOSTDDT-----EIVSY----- 468
Db      2260 ITLEITAVNDAPTLANAIANQATATEDTAFTTIPANTFTDVGADGALTVSATLADGANILP 2319
Qy      469 HWEENGPFEKTSVDSVPLRLSNLDPCGNYSFRLTVTDSGATNSTTAALIVNNA--VD 526
Db      2320 NWLSFNP---STRFTIGTP---TNNSVGTVNIRVATDNAGASVSDVFTLVANSDDTND 2372
Qy      527 YPPVANAGPNH-----TITLPONSITLNGQSDDDHQIVLYEWSLGPSEGKHVVNQ 578
Db      2373 APTLENAIANQATATEDSAFTTIPANTFA-----DVDAGDTLTYSATLADGADLLNWLNF 2427
Qy      579 GVQTPYLHLHSAMQEGDYTFOLKV--TDSSRQOSTAXVTIVOPENNRPPVA-----VAG 630
Db      2428 NPSTRFTSGTPTNDEVTGNTINIKVTATDNAGASLIDFTLTVINTNDAPTVANAIAIANTAT 2487
```

```

Qy      631 PDKELIFPVESATLDGSSSDDHGIVFVWHEVHVGPSAVEMENIDKAIATVTGL----QV 686
Db      2488 EDTAFNQIPA---DAPNDVDTGDTLTYTATLENGDELPSWLTFDAAATRTTSGTPTNSEV 2544
Qy      687 GTYHFRLTVDKQOQGLSSTSTLTVAVKKENSPP-----PRRAGGHRHVLVLPNNSITL 738
Db      2545 DTLSIKVIATDKSQASASNVFTLTVLNTNDAPTLENAIADQTATEDSTFSFIIPVNTFA- 2603
Qy      739 DGSERSTDDQRIVSY-LMIRDQSPAGDVVDGSDHSHVALQLTNLVEGVYTHRLRVTDQSG 797
Db      2604 ----DVDADDILAYSATLEGAALPSWLTFNPTNRTTFAGTPTINSEVGTLNLIKVIATDKSS 2659
Qy      798 ASDTDDTATVEV 808
Db      2660 ANVSDVFTLTV 2670
```

Search completed: October 12, 2005, 10:22:42  
Job time : 58 secs